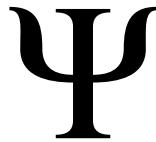**DET PSYKOLOGISKE FAKULTET**

Ψ

An Empirical Investigation of Depression Symptoms:
Norms, Psychometric Characteristics and Factor Structure of the
Beck Depression Inventory-II

HOVEDOPPGAVE

Profesjonsstudiet i psykologi

Harald Aasen

Høst 2001

Forord

Temaet for oppgaven kan bli ubehagelig nærgående når man skriver hovedoppgave om symptomer på depresjon og angst, spesielt når innleveringsfristen nærmer seg. Når dette allikevel ikke har blitt et introspektivt studie, skyldes det først og fremst at 875 tilfeldig oppringte personer og 303 grunnfagsstudenter har brukt av sin dyrebare tid på å fylle ut en rekke relativt nærgående kvantitative selvrapporteringsskalaer om symptomer på depresjon og angst. Veiledere Inger Hilde Nordhus og Ståle Pallesen skal ha en stor takk for å ha initiert prosjektet og gitt meg stor frihet underveis til å utforme min egen oppgave. Videre har Thomas Nordhagen vært en viktig samarbeidspartner etter at han bestemte seg for å legge vekk Kierkegaard og konsentrere seg om psykometrien for en stund. Jørn Hetland og Hilde Mangerud har kommet med nyttige ekspertkommentarer. Til slutt vil jeg takke min kjære Hege som i høyeste grad har sørget for at de mer kvalitative aspektene ved tilværelsen har vært tilstede.

Abstract

The present study provides normative data and investigated the psychometric characteristics and factor structure of a Norwegian version of the Beck Depression Inventory-II (BDI-II) in a general population sample (N=875) and a student sample (N=303). High internal consistency of the BDI-II was found in both samples, and sufficient test-retest reliability over a three week period was obtained in the student sample. Results from the student sample further established evidence for convergent and discriminative validity, but only preliminary indications of the construct validity of the BDI-II were obtained. Confirmational factor analyses were carried out with both samples respectively. Taking into account model parsimony, a three-factor model was retained for post-hoc analyses. A modified three-factor model achieved satisfactory fit across subgroups of the general population sample and the student sample. Sample characteristics and methodological limitations are discussed, validation of the Norwegian version of the BDI-II and contribution to research on the BDI-II in general summarized, and issues for further validation and research recommended.

Introduction

Depression is the second most common psychological disorder, afflicting nearly 100 million people worldwide every year (Beckham & Leber, 1995). Based on data from several epidemiological surveys conducted in the United States, Wittchen, Knauper & Kessler (1994) estimated the lifetime prevalence for major depressive disorder to fall between 15% and 18%. A recent publication from the ODIN study reports a point prevalence of 9% for a large population sample in Norway (Ayuso-Mateos et al., 2001). Several authors are also concerned about the growing prevalence of milder, subclinical forms of depression, and the growing numbers of young people with depressive symptomatology (Barret, Barret, Oxman & Gerber, 1988, Gotlib & Hammen, 1992, Clark & Beck, 1999). Considering these epidemiological findings together with the knowledge about the immense personal, economic and social costs of depression (Simon, Von Korff & Barlow, 1995), it is evident that there is a strong need for an enhanced understanding of this disorder and its symptoms.

In order to meet this need the last decades has brought forth a growing body of research on assessment methods to be used in clinical practice and research on depression (Maruish, 2000, Bech, 1993). One of the most frequently used methods for assessing depression symptoms is the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh 1961, Beck, Steer & Bargin, 1988) and the revised version, the Beck Depression Inventory-II (BDI-II; Beck, Steer & Brown, 1996). A central aim of the present investigation is to contribute to the development of assessment methods to be used in Norway by providing normative data and investigating the psychometric characteristics and factor structure for a Norwegian version of the BDI-II (Nordhus, Pallesen & Aasen, 2001). Additionally, the study is designed to address unsolved issues related to the psychometric characteristics and factor structure of the BDI-II and thereby contribute to the research on the BDI-II in general. Special attention will be given the factor structure since very few studies have employed confirmational factor analysis on the BDI-II, which is now considered the most appropriate methodology for investigating the underlying factor structure of the inventory (Byrne & Baron, 1993, Osman, Downs, Barrios, Kopper, Gutierrez & Chiros, 1998). The study is based on a general population sample (N=875) and a student sample (N=303), and is part of a project on development of assessment instruments in clinical psychology at the Faculty of Psychology, University of Bergen.

The clinical syndrome of depression

One approach to study depression is on the clinical syndrome level. According to advocates of this approach, the experience of clinical depression consists of more than just an emotional experience of being sad or low as a consequence of everyday loss or failure (Gotlib & Hammen, 1992). The clinically depressed person's mood influences, and interacts with, sets of other cognitive, behavioural and bodily functioning symptoms that are described in the current nosological system of psychological disorders, the DSM-IV (American Psychiatric Association, 1994). Although the DSM-IV separates depression into several subtypes, i.e. dysthymic disorder or seasonal affective disorder, most therapists and researchers consider Major Depressive Episode to be prototypical of clinical depression (Clark & Beck, 1999). According to the DSM-IV, the criteria for Major Depressive Episode are that 5 or more of the 9 symptoms in table 1 must be present during the same 2-week period and represent a change from previous functioning:

Table 1: DSM-IV criteria for Major Depressive Episode

Depressed mood	Fatigue or loss of energy
Loss of interest or pleasure	Feelings of worthlessness or guilt
Weight or appetite increase or decrease	Concentration difficulties or indecisiveness
Insomnia or hypersomnia	Suicidal thoughts or ideation
Psychomotoric agitation or retardation	

In addition, at least one of the core symptoms of depressed mood or loss of interest or pleasure must be present and the symptoms must cause significant impairment in social, occupational or other areas of functioning. The diagnosis of major depressive episode would be ruled out if the symptoms meet the criteria for mixed episode of mania and depression, are caused by the effect of a substance, or are better accounted for by bereavement.

The above description from on the DSM-IV gives a picture of the clinical syndrome of depression and the related symptoms. Underlying this diagnostic approach to depression is a traditional medical model of psychopathology, which assumes a categorical perspective on depression. According to a categorical perspective on depression, clinical or diagnosable

depression states as described previously, are qualitatively different from subtreshhold depression and non-clinical states (Coyne, 1994, Gotlib & Hammen, 1992). Often referred to as support for this perspective, several studies have shown that moderate or high scores on self-report measures of depression symptoms are not necessarily indicative of clinical depression when followed up with a diagnostic interview (Ennis, Barnes, Kennedy & Trachtenberg, 1998, Boyd, Weissman, Thompson & Myers, 1982, Myers & Weissman, 1980).

On the other hand, the categorical perspective on psychopathology in general, and depression in particular, has been criticized for the lack of validity of the categories in the diagnostic systems employed (Kline, 2000, Clark & Beck, 1999, Clark, Watson & Reynolds, 1995, Costello, 1993). There is, for example, strong evidence for the comorbidity between the several of the categories in the DSM-IV manual (Clark & Beck, 1999, Clark et al., 1995). In the case of depression, researchers have found a high comorbidity rates with anxiety (Maser & Cloeninger, 1990) and personality disorders (Farmer & Nelson-Gray, 1990). As a result of this critique, several authors (i.e. Clark & Beck, 1999, Costello, 1993) have advocated a dimensional perspective on depression with focus on depression symptoms.

Depression symptoms - a dimensional perspective

The empirical focus of the present investigation is depression symptoms in non-clinical subjects measured by a dimensional measure, and not the clinical syndrome as described in the DSM-IV. According to a dimensional perspective underlying this approach, depression is a continuous, quantitative dimension. An approach on the symptom level is therefore viewed as relevant and important to the understanding of depression (Clark & Beck, 1999). Opposing to a categorical perspective, Clark & Beck (1999) argue that clinical and non-clinical depression states vary only in the severity of symptoms, not in kind. Different types of evidence are referred to as support for their position. First, several empirical studies have found that individuals with few depression symptoms are at higher risk for developing a subsequent major depressive episode (Flett, Vredenburg & Krames, 1997). Flett et al. (1997) refer to this as etiologic continuity and conclude in their review on the topic that it has substantial support in the literature. Second, presence of a few depressive symptoms have also been associated with considerable functional impairment that differ from major depression only in degree rather than kind. For example, in a longitudinal study, Gotlib

(1985) found that subjects with elevated self-report depression scores but no diagnosable depression, did not differ significantly from individuals with diagnosable depression on most variables on psychosocial functioning. Third, several studies (i.e. Fechner-Bates et al., 1994, Weissmann, Prusoff & Pincus, 1975) have found more similarities than differences in the symptoms reported by those with clinical and non-clinical depressive states. Based on this empirical evidence, it can be argued in line with Clark & Beck (1999) that individuals with diagnosable depression will have a greater number of depressive symptoms at higher levels of severity than individuals with subthreshold depressive symptoms, but no qualitatively different symptom characteristics can be identified that distinguish major depression from subtreshold depressive symptoms or non-clinical states.

The dimensional/categorical debate in depression research continues to be contentious in depression research, and Flett et al. (1997) have suggested a more moderate position on the issue. They conclude in their review of the topic that even though there is mounting evidence for the dimensional perspective, researchers should remain open to the possibility that there are certain aspects of depression that are discontinuous and categorical (Flett et al., 1997).

Implications of the perspectives for research on depression

Most important regarding the present investigation are the implications the two perspectives have for research and assessment of depression. A categorical perspective will regard investigation into depression symptoms of non-clinical subjects to be of questionable relevance and generalizability to the understanding of clinical depression states. Gotlib & Hammen (1992), for example, conclude that although depression may be scaled on a single dimension, qualitative differences exist that distinguish more severe, diagnosable depression. They also suggest that low-grade depressive symptoms in non-clinical subjects may reflect negative affectivity (Watson & Clark, 1984) or general psychological distress (Gotlib, 1984), and not depression as such.

According to the dimensional perspective, however, depressive symptoms in persons drawn from non-clinical populations are postulated to represent the low end of a continuum of severity - with clinical depression as the opposite pole, quantitatively, but not qualitatively different. Research on depressive symptoms in non-clinical samples is therefore seen as relevant and important to the understanding of the disorder (Clark & Beck, 1999).

Self-report scales of depression symptoms

A type of objective assessment method that has become increasingly popular the last decades is self-report scales. Self-report scales are relatively brief, quantitative instruments constructed to assess the presence and severity of psychopathology dimensions and not for diagnostic purposes (Maruish, 1994). Further in line with a dimensional perspective on depression symptoms, most self-report scales are designed to quantitatively measure the severity of depression symptoms on a continuum.

Nezu, Ronan, Meadows & McClure (2000) have found over 30 self-report scales of depression symptoms reported in the literature, and several of these have achieved widespread use. Derogatis & Dellapietra (1994) have outlined some of the reasons why self-report measures of depression have become frequently used. First, from practical point of view, they are time- and cost efficient, easily administered and scored, and they and can be used in a variety of settings. This is probably why many authors believe they can help solving the problems of increasing costs in mental health care and contribute to more efficient services (Maruish, 2000). Second, they avoid inter-rater inconsistence that can be a problem in interviewing, an attribute also that make them suitable for research purposes. Third, the majority of the self-report scales of depression have shown to have high internal consistency, which is of major importance both in clinical practice and research (Derogatis & Dellapietra, 1994). A final advantage that should be mentioned is that self-report scales are completed by the only person experiencing the symptoms – the client, and thus give an important indication of the subjective experience of the symptoms that otherwise may be difficult to assess.

This latter advantage is of course also one of the most important disadvantages of self-report scales of depression. The clients' bias and defensiveness in reporting their symptoms may be a source of error (Joiner, Schmidt & Metalsky, 1994). This is especially potent if there are personal gains associated with response distortion (Joiner, Schmidt, Cook & Gencoz, 2000). A closely related disadvantage is that most self-report scales are relatively face valid. This attribute certainly makes it easier for the respondent to 'fake good or bad' depending on the personal motivation for completing the scale. Finally, although advocates of a dimensional perspective on depression symptoms regard self-report scales of depression as relatively sensitive (Clark & Beck, 1999), several authors argue that they are not necessarily very specific of depression (Katz, Shaw, Vallis & Kasier, 1995, Coyne 1994). As mentioned, high scores self-report scales of depression symptoms are not always correlated

with the clinical diagnosis of depression (Ennis, Barnes, Kennedy & Trachtenberg, 1998) and according to a categorical perspective on depression this supports the interpretation of scores on self-report scales on depression symptoms as not necessarily indicative of clinical depression (Gotlib & Hammen, 1992, Coyne, 1994).

The Beck Depression Inventory

Development

The self-report scale in focus in this study is the Beck Depression Inventory-II (BDI-II, Beck et al. 1996). The BDI-II is a quantitative, self-report scale for measuring the presence and severity of depression in clinical and normal populations of adults and adolescents. The original BDI (Beck et al. 1961) was empirically constructed based on a pool of pre-selected items from Beck's observations in psychotherapy with depressed patients. According to Beck (1967) 'the items were chosen on the basis of their relationship to the overt behavioural manifestations of depression and do not reflect any theory regarding the etiology or the underlying psychological processes in depression' (p. 189). The scale was then constructed through discriminative analysis of these items regarding symptoms that differentiated depressed from non-depressed. The first version of the BDI ended up consisting of 21 items concerning symptoms and attitudes that appeared to be specific for depressed patients: depressed mood, pessimism, failure, satisfaction, guilt, punishment feelings, self-dislike, self-accusations, suicidality, crying, irritability, social withdrawal, indecisiveness, body image, work difficulty, sleep, fatigue, appetite, weight, somatisation and loss of libido (Beck et al, 1961).

The original version of the BDI has been revised twice since the first version was published. In 1979 (Beck, Rush, Shaw, & Emery, 1979) eliminated alternative wordings for the same symptoms and avoided double negatives. The number of response options was limited to four alternatives, and minimal wording modifications were made for 15 items. A technical manual for this revised version (BDI-IA) was published in 1987 (Beck & Steer, 1987). Beck et al. (1996) note that despite the publication of the BDI-IA, the original 1961-reference remained the most frequently cited.

In the same period, Moran and Lambert (1983), and later Vredenburg, Krames and Flett (1985), questioned whether the BDI and the BDI-IA adequately addressed the newly

released DSM-III criteria for major depressive episode (Diagnostic and Statistical Manual on Mental Disorders - 3 edition, 1980). The consecutive releases of DSM-III-R and DSM-IV resulted in a new revision of the BDI to the BDI-II, aiming to assess symptoms that were more consonant with the manuals. Four items, concerning body image, work difficulty, weight loss and somatic preoccupation were left out and replaced by new items concerning agitation, worthlessness, loss of energy and concentration difficulty. The insomnia and loss of appetite items were reworded to reflect both increases and decreases. Several of the statements used in the rating of the items were reworded, and the time frame for the BDI-II was extended to two weeks instead of the 'past week, including today' in the BDI.

Table 2: BDI-II items

Sadness	Self-dislike	Indecisiveness
Pessimism	Self-criticalness	Worthlessness
Past failure	Suicidal thoughts or wishes	Loss of energy
Loss of pleasure	Crying	Sleep changes
Guilt feelings	Agitation	Irritability
Punishment feelings	Loss of interest	Appetite changes
Concentration	Tiredness	Loss of sexual interest

Beck and colleagues then conducted a series of item and factor analyses with psychiatric outpatients and university students, and published the BDI-II manual in 1996 (Beck, Steer & Brown, 1996). The BDI-II items are listed in table 2 previously, and a closer description of the measure is given in the methods section.

The BDI-II and perspectives on assessment of depression

Although Beck (1967) has argued that the BDI is atheoretical in nature, there is agreement in modern psychometric literature that any assessment method must be based upon theoretical considerations about the construct being measures (Kline, 2000). Seen in the light of the perspectives on depression outlined previously, several theoretical considerations underlying the use of the BDI-II should be mentioned. First, based on a dimensional

perspective on depression, Beck and colleagues (Beck et al. 1988, Clark & Beck, 1999) assume that scores on the BDI is both relevant and important to the understanding of depression. This also applies to BDI scores obtained in non-clinical samples, which is the focus in the present investigation. Second, in line with a quantitative assessment perspective, it is assumed that standardized and well-validated self-report scales provide relevant information about the respondents' depression symptoms. Thus, it is assumed that depression symptoms like i.e. guilt and self-dislike can be quantified, measured on a continuum and interpreted as meaningful information about the clients subjective experience. And finally, it is important to note that although the BDI have been revised in order to be more consonant with the DSM-IV criteria for Major Depressive Disorder, inspection of table 1 and 2 reveals that the revised version of the inventory still diverges from the DSM-IV in that it has far more subjective items than somatic items. According to Clark & Beck (1999) the relative strong emphasis on somatic criteria in the DSM-IV is not supported by empirical evidence, and in line with a cognitive model of depression (Beck, 1967, Beck et al., 1979) the BDI-II has included more subjective items than somatic.

These theoretical influences on the BDI are not tested directly in the present psychometric investigation. They nevertheless constitute a theoretical framework for the inventory, which has important implications for the interpretation of the scale.

Review of research on the BDI-II

The original BDI has been reviewed extensively with respect to psychometric properties and factor structure (Beck et al., 1988). According to Beck et al. (1996) the BDI-II constituted a 'substantial revision' of the original BDI. This review of the descriptive data, psychometric characteristics and factor structure of the BDI-II will therefore only briefly mention results from research on the original BDI when no such research has been conducted on the BDI-II. Further, since this investigation is based on two non-clinical samples, the review of earlier studies will focus on studies with non-clinical samples, and only exceptionally mention studies based on clinical groups.

Standardization and normative comparisons

The advantages of self-report scales outlined previously are based on the assumption that the scale is standardized for use in a defined population. This implies that the scale must be administered to a representative sample of the target population, and examined with respect to descriptive data and psychometric characteristics (Cicchetti, 1994, Cohen, Swerdlik & Phillips, 1996). For clinical scales, standardization of the scale also implies that descriptive data from non-clinical populations are examined in order to have normative comparison for clinical populations (Kendall & Sheldrick, 2000).

Very few clinical self-report scales currently in use in Norway are translated, standardized and examined with respect to psychometric properties with a Norwegian general population sample (Vassend, Lian & Andersen, 1992). The present study will provide this for the BDI-II, and thus contribute to the development of clinical assessment methods to be used in clinical practice and research in Norway.

Furthermore, the majority of assessment methods in clinical psychology are relying on student samples as non-clinical normative comparisons, and the BDI is no exception. Beck et al. (1988) reviewed 18 studies of normal populations data on the original BDI, and none of these were based on general population samples. More recently, Kendall & Sheldrick (2000) reviewed the normative basis for measurements of depression symptoms published in the Journal of Clinical Psychology, and found 9 studies on the BDI, all based on student samples. The only exceptions we have found in the literature on the BDI, is the recent European ODIN study (ODIN study report, 2001) where the original version of the BDI was used to screen for depression in the general population in several European countries. However, this study did not provide any normative data.

Student samples are not representative of the general population on several demographic variables like sex, education and marital status, and are often referred to as 'convenient samples' (Enns, Cox & Borger, 2000). Kendall & Sheldrick (2000) therefore emphasize the importance of obtaining normative data from a representative sample of the target population. Also, as far as the BDI-II is concerned, Beck et al. (1996), and Steer & Clark (1997) have recommended that the normative basis and the psychometric characteristics of the scale must be investigated in a variety of non-clinical samples, not just in university students. To date, no studies have employed the BDI-II in a general population sample, and the present study brings the research on the BDI-II further by providing data from a general population sample from the Norwegian population.

Descriptive data on the BDI-II

In the BDI-II manual, Beck et al. (1996) provide descriptive data from a student sample (N=120). The total mean of the scale was 12.6 (SD=9.9), which is indicating minimal levels of symptomatology according to the scoring criteria in the manual. The item means were ranging from .26 (loss of interest in sex) to .97 (changes in sleep). According to Beck et al. (1996) each individual item contributed significantly to the total score, ranging from a correlation of .27 for item 21 (loss of sexual interest) to .74 for item 7 (self-dislike). They found a significant sex difference with women reporting higher total scale scores than men (14.6 vs. 10.0), and age was significantly correlated with the BDI-II score in this sample (r=-.18, p<.001).

As can be seen in table 3 below, another five studies have investigated the descriptive and psychometric characteristics of the BDI-II with non-clinical student samples after the publication of the BDI-II manual. The means of the total BDI-II-score in these studies are ranging from 8.4 (SD=7.2) in the study by Whisman, Perez & Ramel (2000) to 11.9 (SD=8.1) reported by Steer & Clarke (1997). All of the means are within the 'minimal' symptom group as described in the manual (Beck et al., 1996). Also the individual item means seem to differ across samples. For example, Osman, Downs, Barrios, Kopper, Gutierrez & Chiros (1998) report that 4 of their item means differentiated their sample from Beck et al. (1996) sample.

Insert table 3 about here

Examination of item-total correlations indicates that item 7 (self-dislike) is consistently associated with the total scale score in non-clinical samples, while item 21 (loss of sexual interest) correlates less with the total scale score (Beck et a. 1996, Osman et al., 1998, Whisman et al., 2000).

As far as sex differences are concerned, the results are inconsistent. Osman et al. (1997) reports in line with Beck et al. (1996) a significant sex difference with women scoring higher than men, while three of the other studies reports no significant difference. Taking further into account the earlier contradictory results on sex differences, O'hara, Sprinkle & Ricci

(1998) conclude that additional research investigating sex differences in mean total BDI-II scores is needed.

Psychometric characteristics

Reliability. The internal consistency of the original BDI has been reviewed thoroughly by Beck et al. (1988). In their meta-analysis they report a mean coefficient alpha of .81 based on 24 studies with non-clinical subjects (mostly university students). They also find consistently high coefficient alphas across different clinical samples.

The internal consistency of the BDI-II is reported to be higher than for the original BDI. In the BDI-II manual, Beck et al. (1996) report a coefficient alpha of .93. This result is further supported by the other recent studies on the BDI-II. The coefficient alpha values found are ranging from .89 (Whisman et al, 2000, Steer & Clarke, 1997) to .91 (Dozois, Dobson & Ahnberg, 1998), which must be considered as consistently sufficient. Although the internal consistency of the BDI-II appears in these studies to be strong, Steer & Clarke (1997) recommend further investigation of the internal consistency of the BDI-II across different samples.

Test-retest reliability. As far as test-retest reliability is concerned, Beck et al (1988) report test-retest estimates ranging from .62 (4 months) to .90 (2 weeks period) in their review on the original BDI. For example, from a test-retest study on the original BDI with an interval of 3 weeks, the authors report an estimate of .78 (Oliver & Burkham, 1979).

The only test-retest study on the BDI-II to date is Beck et al. (1996) study of a subsample including only 26 psychiatric outpatients reported in the manual. They found a one-week test-retest correlation of .93. Taking into account that the specification of time span is extended to two weeks in the BDI-II, and the limited generalizability of Beck et al. (1996) re-test, there is a strong need for further exploration of the test-retest stability of the BDI-II in non-clinical samples. The present study will extend the research on the BDI-II by investigating the test-retest of the BDI-II in a student sample over a three-week period.

Convergent validity. Despite the revision the BDI, one should expect a relatively high correlation between the BDI and the BDI-II. In line with this expectation, Beck et al. (1996) reports in the manual a high correlation between the original BDI and the BDI-II of .93 with their clinical sample. This finding is replicated with a student sample by Dozois et al. (1998) (r=.93). Earlier research on the original BDI with non-clinical samples reveals consistently

high correlations (mean r=.71) with the Zung-SDS (Zung, 1965)(Beck et al, 1988) and the SCL-90 depression subscale (SCL-90; Symptom Check List 90, Derogatis, 1983)(r=.71, Rouseville, 1979, cited in Beck et al., 1988, and r=.68, Gotlib, 1984). Furthermore, Osman et al. (1998) report in their recent study that the BDI-II correlates substantially (r=.77) with the DASS-depression scale (DASS; Depression Anxiety Stress Scale, Lovinbond & Lovibond, 1993), thus providing evidence for the convergent validity of the BDI-II.

No other studies than the one by Osman et al. (1998) have investigated the convergent validity of the BDI-II with other measures of depression symptoms. In order to investigate the convergent validity of the BDI-II further, the Zung-SDS and SCL-90 depression subscale is employed as convergent measures in the present investigation. The Zung-SDS scale has shown evidence for the convergent validity with the original BDI (Beck et al., 1988), which should be further investigated on the BDI-II. The SCL-90 depression subscale is employed in order to investigate the convergent validity of the BDI-II with one of the few self-report symptom scales that are already translated and validated for use in Norway (Vassend et al., 1992).

Discriminant validity. The concept of discriminant validity refers here to the low association between the BDI-II and a measure that is supposed to measure a different construct. Accordingly, the correlation between the BDI-II and the discriminant measure is expected to be significantly lower than with a convergent measure.

There is some disagreement in the literature on the relationship between the symptoms of depression and anxiety. According to a cognitive model of the disorders of depression and anxiety, the two disorders can be separated by its cognitive content, that is loss for depression and fear for anxiety (Beck, 1967, Beck et al.,1979). Despite this theoretical separation, there is substantial evidence for the comorbidity of the two disorders (Maser & Cloninger, 1990). Also on a symptom level, researchers have found it difficult to differentiate the two constructs, especially in non-clinical samples (Gotlib, 1984, Joiner, 1996). Advocates of a categorical perspective on depression have therefore referred to this as evidence for a more general construct like negative affectivity (Watson & Clark, 1984) or general distress (Gotlib, 1984). In research on the BDI, however, the Beck Anxiety Inventory (Beck & Steer, 1993a) is often used as a discriminant measure (Beck et al., 1988).

In the BDI-II manual, Beck et al. (1996) reports a correlation of .60 between the BDI-II and BAI, and a correlation of .47 between the BDI-II and the Hamilton Anxiey Rating Scale (Hamilton, 1959), for the their outpatient sample. They conclude that these results indicate

robust discriminant validity between depression and anxiety. Osman et al. (1997) refer to this conclusion and question whether the moderately high correlation of .60 between BDI-II and BAI actually suggests strong evidence for discriminant validity of the BDI-II. In their study with a student sample they find a correlation of .56 between the BDI-II and the BAI, which is identical to Steer & Clarke (1997) result with their student sample. Interestingly, Steer & Clarke (1997) report the correlation of .56 between the BDI-II and BAI as evidence for convergent validity of the BDI-II. These diverging reports probably reflect the disagreement in the literature on the division of depression and anxiety symptoms. In the present investigation the BAI is employed in order to further investigate the validity of the BDI-II.

Theoretically, the SCL-90 psychotisism subscale should measure symptoms that are easier to discriminate from depression symptoms than anxiety symptoms. According to Derogatis (1983) the SCL-90 psychotisism subscale measures a continuum of psychotic symptoms ranging from schizoid traits to hallucinations and delusions. However, in an earlier study by Gotlib (1984) with a student sample, a correlation coefficient of .59 between the SCL-90 psychotisism subscale and the original BDI is reported, which must be considered as moderate. No studies to date have employed the BDI-II and the SCL-90 psychotisism subscale. In the present study, the discriminant validity of the BDI-II is investigated further by employing the SCL-90 psychotisism subscale in the student sample.

Construct validity. Osman et al. (1998) have provided some evidence for the construct validity of the BDI-II in relation to self-esteem and stress. In their study they report a correlation of -.60 with Rosenberg Self Esteem scale (RSE; Rosenberg, 1979) and .68 with the DASS-stress scale (DASS; Lovinbond & Lovibond, 1993). This is the only study to date that has investigated the construct validity of the BDI-II, and several authors argue that the construct validity of the BDI-II must be investigated with respect to a variety of measures (Steer & Clark, 1997, Dozois et al, 1998).

According to the reformulated hopelessness theory of depression, attributional style predicts depression (Abramson, Seligman & Teasdale, 1978). More specifically, the theory postulates that internal, stable and global attributions that are offered for positive and negative events are associated with depressive symptoms. Several studies have shown that both clinical and non-clinical subject's responses to the Attributional Style Questionnaire (ASQ) are associated with self-rating scales of depression symptoms (Peterson, Semmel, von Baeyer, Abramson, Metalsky, Seligman (1982). This is also supported by research on a

Norwegian version of the ASQ by Hoffart and colleagues in several clinical samples (Hoffart & Martinsen, 1991, Hoffart & Torgersen, 1991).

A further investigation of the association between the BDI-II and the ASQ can add more insight into the construct validity of the BDI-II. The ASQ is therefore employed as a measure in the student sample in our study in order to investigate the construct validity of the BDI-II.

Factor structure of the BDI-II

An important issue in research on depression symptoms in general, and this investigation particular, is the weighting of the subjective symptoms versus somatic symptoms of depression. As mentioned earlier, the BDI-II emphasizes subjective symptoms to the relative neglect of somatic symptoms. Altogether 16 of the items (sadness, pessimism, failure, guilt, punishment, self-dislike, self-criticalness, suicidal thoughts, crying, worthlessness, loss of pleasure, loss of interest, indecisiveness, loss of energy, irritability and concentration) may be viewed as subjective items, compared to only 5 items (sex, appetite, fatigue, sleep, and agitation) referring to somatic symptoms. Thus, although the BDI was revised in order to be more consonant with the DSM-IV criteria, the weighting of subjective versus somatic/motivational items is the opposite of the DSM-IV, and the inventory has several subjective items not mentioned in the manual.

Factor analytic studies of the BDI-II studies have showed that it is possible to empirically separate the BDI-II items into two underlying factors that more or less correspond to the separation of subjective and somatic items outlined above. Beck et al. (1996) performed exploratory factor analyses on their student sample reported in the manual and extracted two correlating factors: a cognitive-affective factor consisting of the subjective items sadness, failure, loss of pleasure, guilt, punishment, self-dislike, self criticalness, suicidal thoughts, crying, agitation, loss of interest, indecisiveness, worthlessness and irritability, and a somatic factor consisting of items loss of energy, sleep, appetite, concentration and fatigue (correlation between factors r=.62). The items 2 (pessimism) and 21 (loss of interest in sex) did not load significantly (<.35) to any of the factors. The results of this factor analysis were replicated by Steer & Clark (1997) employing the same explorative methodology. In addition, these researchers found that the items 2 and 21 loaded on the cognitive-affective dimension.

Osman et al. (1998) comment that these two studies on the BDI-II use explorative factor analyses (EFA) and not confirmational factor analyses (CFA), which is now agreed upon as a better method for investigating the factor structure of well-established self-report measures of depression (Byrne, 2001, Byrne & Baron, 1993, Osman et al., 1998, Dozois et al., 1998). Most of the recent studies validating the factor structure of the original BDI have used CFA with Structural Equation Modelling techniques (Byrne & Baron, 1993, Byrne, Baron, Larsson & Mehlin, 1995). The advantage of CFA of a scale like the BDI-II is that researchers are allowed to test the fit of a predefined model based on theory and empirical research on the data, instead of a data-driven procedure in EFA.

Only three studies have employed CFA to investigate the factor structure of the BDI-II with non-clinical samples (Osman et al., 1998, Dozois et al., 1998, Whisman et al., 2000). In their study, Osman et al. (1998) applied CFA in order to evaluate the fit of the oblique two-factor model as reported in the manual (Beck et al., 1996), compared to the fit of a modified three-factor model (previously found by Byrne & Baron (1993) on the original BDI) and a one-factor model as alternative models. The three-factor model provided best fit to the data, while the hypothesized to-factor model only met two of the pre-established criteria. Osman et al. (1998) therefore concluded that the BDI-II may tap three dimensions of depression severity in a non-clinical sample. In line with Byrne & Baron (1993a) earlier research on the original BDI, they called the three factors 'negative attitude', 'performance difficulty' and 'somatic elements'. Contrary to the earlier studies, item 21 (loss of sexual interest) was constrained to load on the somatic factor in this three-factor model. The authors also noted, however, that the two-factor model can be improved by allowing item 19 (concentration) to load on the cognitive-affective factor and correlated errors between items 6 and 8, 10 and 16 and 15 and 20.

Dozois et al. (1998) employed both EFA and CFA with their large student sample (N=1022). They divided the sample and employed EFA on the first half, and CFA on the second. In concordance with factor analysis reported in the manual (Beck et al, 1996) they reported two factors from the EFA, a cognitive-affective factor accounting for 38% of the variance, and a somatic factor accounting for 8% of the variance. The correlation between the factors was .60. In the following CFA, the two-factor solution produced good fit. Also a recent study by Whisman et al. (2000) further supports a two-factor solution. They performed a CFA on their student sample, and when allowing item 2 (pessimism) and 21 (loss interest in sex) to load on the cognitive-affective factor, and several residuals correlate

(item 1 and 10, item 4 and 12, and item 7 and 8), the model achieved satisfactory fit. Again, the factors correlated substantially (r=.71).

Earlier studies by Byrne & Baron (1993a) and Byrne, Baron, Larsson & Mehlin (1995) deserve attention although the original version of the BDI was employed. These authors note that the factors found on the BDI is often highly correlated, suggesting the presence of a second-order general factor (Byrne et al., 1995). In their studies on non-clinical Canadian and Swedish adolescents (Byrne & Baron, 1993, Byrne et al., 1995), they found good fit for hypothesized second-order model with the general factor 'depression' as a second-order factor and the three factors mentioned in the study by Osman et al. (1998) as first-order factors.

Overall, the results from these studies on the factor structure of the BDI-II provide considerable support for a modified oblique two-factor model where item 2 and 21 are constrained to load on the cognitive-affective factor. Osman et al. (1998) provide evidence for an oblique three-factor model where item 21 (loss of sexual interest) is loading on the somatic factor, but they also note that a two-factor model produce acceptable fit if improved. A second-order model, which has shown good fit for the original BDI, has never been tested with the BDI-II. Since most of the studies on the factor structure of the BDI-II report substantial correlation between the factors, a second-order factor model should also be tested out on the BDI-II.

Aims and expectations of the present investigation

The present investigation of depression symptoms measured by the BDI-II has two related aims. First, it is aimed to provide normative data and investigate the psychometric characteristics and factor structure of a Norwegian BDI-II. Second, it is aimed to meet the need for further research related to the issues reviewed previously, with special attention to the factor structure of the BDI-II.

The specific aims and expectations of the study can now be summarized as follows. First, descriptive and correlational analyses of the BDI-II scales and items are carried out with the general population sample and the student sample respectively and the results are expected to be comparable to previous results on the BDI-II with non-clinical samples.

Second, psychometric characteristics and factor structure on the BDI-II is examined. The internal reliability of the BDI-II is expected to be high in both samples, and the test-retest is expected to further confirm the reliability of the scale in the student sample. Convergent, discriminant and construct validity is examined in the student sample, and the BDI-II is expected to correlate high with the Zung-SDS and the SCL-90 depression subscale, moderately with the BAI, and significantly lower with the SCL-90 psychotisism subscale than the correlation with the convergent measures. With respect to construct validity, total BDI-II scores above 13 are expected to be associated with increasing internality, stability and globality attributions for negative events, and an opposite pattern for positive events.

Finally, confirmatory factor analyses will be carried out in both samples in order to test the fit of the following hypothesized factor models based on recent research: a two-factor model, a three-factor model and a second-order three-factor model (a closer description of the models is given in the results section later). The retained factor model for both samples will be modified in post-hoc modification analyses (Jøreskog, 1993, Byrne, 2001) in order to obtain a better fitting model for the general population sample and student sample respectively. The modified model for the general population sample will then will be tested on subgroups by sex, education and marital status.

Methods

Measures

Demographic data. Demographic data about sex, age, education and marital status were obtained in both samples.

The Beck Depression Inventory-II. The Norwegian version of the Beck Depression Inventory-II (Beck et al., 1996, Nordhus et al., 2001) is a 21-item self-report instrument for measuring the presence and severity of depression symptoms in clinical and non-clinical adults and adolescents over 13 years old. Respondents are asked to pick out the corresponding number to the one out of four statements that best describes the way they have

been feeling during the past two weeks, including today. The items are presented in table 2 previously. Each item is rated on a 4-point scale (0-3) in increasing severity, adding up to a total score range from 0 to 63. Reported reliability and validity of the scale is reviewed previously.

The Beck Anxiety Inventory. The Norwegian version of the Beck Anxiety Inventory (BAI, Beck & Steer, 1993a, Nordhus, Pallesen & Nordhagen, 2001) is a self-report measure of the severity of anxiety symptoms experienced by the respondents the last two weeks, including today. The items are drawn from earlier self-report instruments that measure various aspects of anxiety, like the Anxiety Check List (ACL, Beck, Steer & Brown, 1985) and the Situational Anxiety Check List (SAC, Beck, 1982). Each of the 21 descriptive items is rated on a 4-point scale from 'not at all' to 'severely'. The original BAI have shown sufficient reliability and convergent and discriminant validity (Beck & Steer, 1993a), which is further supported by validation of the Norwegian version (Nordhagen, 2001).

The Norwegian versions of the BDI-II and the BAI were constructed employing a back translation procedure as described by Geisinger (1994). The instruments were translated from English to Norwegian by the authors, and back to English by a professional bilingual translator not familiar with the instruments. This second translation was compared with the original versions of the instruments and it was ensured that the original meaning of the items was maintained in the translated versions.

Symptom Check List 90-R. Norwegian versions of the depression and psychotisism subscales of the revised Symptom Check List 90 (Vassend et al., 1992, Derogatis, 1983) were employed as convergent and discriminant measures respectively in the student sample. The depression subscale consists of 13 items measuring depression symptoms the respondents have experienced the last week. Each item is rated on a scale from 0 (not at all) to 4 (very much). The psychotisism scale is constructed similarly, consisting of 10 items measuring psychotic symptoms like 'hearing voices' and 'feeling someone can read your mind'. The items of these two subscales were randomly presented in the questionnaire. Vassend et al. (1992) reports satisfying internal reliability for both subscales for the Norwegian version of the revised SCL-90, adding to the reported reliability and validity of the original SCL-90 reported by Derogatis (1983). Reliability analysis of our data produces satisfactory alpha values of .89 for the depression subscale, and .71 for the psychotisism subscale (table 13, appendix).

The Zung Self-Rating Depression Scale. The Zung-SDS (Zung, 1965) is a frequently used depression scale consisting of 20 items assessing symptoms of depression identified in factor analytic studies of depression symptoms. Ten of the items are worded symptomatically positive, and ten symptomatically negative. The respondents are asked to indicate how often each statement is a correct description of their state the last week on a 4-point scale from 'none of the time' to 'all of the time'. The symptomatically positive symptoms are graded from 1 to 4, while the symptomatically negative items are graded from 4 to 1. Authors report alphas of .88 to .93 for different groups (Nezu et al., 2000). A Norwegian version of the Zung-SDS was constructed by translation of the original version by the author and applied as a convergent measure in the student sample in this study. The scale produced sufficient reliability (alpha=.85, see table 13, appendix)

The Attributional Style Questionnaire. The Attributional Style Questionnaire (ASQ; Peterson et al., 1982) is a self-report measure used to assess respondents' causal judgements of 6 positive and 6 negative events. The events described relates to either affiliation (e.g. friend offer a compliment) or achievement (e.g. becoming rich). After stating the most likely cause, each event is rated on a 7-point scale for the dimensions of internality, stability, globality and importance. High scores are associated with increasing ratings of these dimensions. Sufficient reliability and validity for the scale has been reported (Peterson et al., 1982), and a Norwegian version has been applied in clinical subsamples of depression (Hoffart & Martinsen, 1991) and agoraphobic patients (Hoffart & Torgersen, 1991). Reliability measures for the Norwegian version was comparable to those found in the original studies of the ASQ, ranging from alpha= .56 to .67 for the positive events, and alpha=.62 to .76 for the negative events (Hoffart & Martinsen, 1991). In the present study a revised version of the Norwegian ASQ (Hoffart, 1985) was applied as a convergent measure in the student sample. As can be seen in table 13 in the appendix the alpha values for the positive events were .49, .61, .67 and .67 for internality, stability, globality and importance respectively. The corresponding alpha values for the negative events were .47, .64, 75 and .70. These alpha values were on average comparable to those of Hoffart & Martinsen (1991), but the low alpha values of the internality dimensions indicate low internal reliability of this subscale.

Samples and procedure

General population sample. A sample of Norwegian population (N=875, 512 women and 363 men) over 18 years was recruited by Opinion Research Institute October and November 2000. Respondents were randomly recruited from the Telenor register of private numbers. A total number of 4738 private numbers were called always asking for the person who last had birthday in order to ensure that each subject in the household had an equal chance of being asked to participate in the survey. Out of the total number called, 1326 did not answer after 6 calls. Altogether 1409 subjects refused to participate. Out of the remaining 2003 subjects who agreed to be sent the the subsequent letter, 875 returned completed questionnaires, resulting in a response rate of 43,7% for the normative sample. The participants were given a short description of the study in the letter and completed the BDI-II and BAI together with two other self-report scales not described in this study.

The mean age of the sample was 45,4 years for men (range 18-86) and 43,5 years for women (range 18-86). Altogether 49,8% of the subjects had college or university education, and 75,3% were married or living with their partner. The remaining 24,6% were either single, divorced or widow/widower.

Insert table 4 about here

As can be seen from table 4 the randomly selected subjects were not evenly distributed by age and sex. In order to obtain representative data of the Norwegian population over 18 years of age, the data were weighted (for details on procedure see Kessler, McGonagh, Zhao, Nelson, Hughes, Eshleman, Wittchen & Kendler, 1994) according to the distribution of age and sex in the Norwegian adult population (Statistical Central Bureau, 2001). The weights were provided by Opinion Research Institute (2001).

Student sample. The student sample (N=303) consisting of 212 women and 91 men, were obtained at psychology undergraduate lectures at the University of Bergen, April and September 2001. The subjects were given the questionnaires and a brief explanation of the study, and kindly asked to obtain the number of their questionnaires for the re-test. The students completed the scales in following order: BDI-II, BAI, Zung-SDS, SCL-90-depression and psychotisism subscales and the ASQ. Altogether 54 students did not

complete the more time demanding ASQ due to shortage of time. There was no significant difference in mean total BDI-II score of those not completing the ASQ compared to those who did (t(1, 301)=1.5, p=.124).

The mean age of the students was 21.9 (range 19-38) for men and 20.9 (range 18-45) for women. The majority of the students (82,8%) were single, and 15,3% were partners or married.

Re-test sample. The BDI-II and the BAI were re-tested three weeks after the first administration. A total of 128 students (99 women and 29 men) completed the questionnaires at both administrations, yielding a response rate for the test-retest of 42,2%. There were a significantly higher proportion of women in the re-test group compared with the group not participating in the re-test (χ^2(1, 301)=4.8, p>.028). No other significant demographic differences (age: t(1, 301)=.44, p>.672, marital status: (χ^2(1, 301)=0.11, p>.737), nor significant differences in total BDI-II score at the first administration (t(1, 301)=-1.26, p>.209), were found between the two groups.

Statistical procedures

Analyses of descriptive and correlational data, and scale reliability and validity analyses were performed in SPSS.10 and Statistica 6. All total scale scores were computed by adding up individual item scores. For the positive and negative events in the ASQ, total scale scores and alpha values were estimated for each attribution dimension separately. The confirmational factor analyses were carried out with Structural Equation Modelling in AMOS 4. (Arbuckle, 1995). Following fit indices were applied in the analyses: Chi-square, CFI (Comparative Fit Index, Bentler, 1990), AIC (Akaike Information Criterion, Akaike, 1987) and RMSEA (Root Mean Square Error of Approximation, Browne & Kudec, 1993). Applying the criteria used by Osman et al. (1998) and Dozois et al. (1998), the CFI should be higher than .90, and the RMSEA lower than .10 to indicate satisfactory fit. According to Byrne (2001) the chi-square is sensitive to sample size, which make the index problematic to interpret in the present investigation. Since the models tested are not hierarchical, the selection of the retained model for the post-hoc analyses will be based on the AIC index, with the lowest value indicating best fit taking into account model parsimony (Schumacker & Lomax, 1996). The retained models will be modified based on the modification indices feature in Amos 4. and earlier research. These post-hoc analyses are not strictly

confirmational, but are more in line with what Jøreskog (1993) calls model generating procedures (see also Byrne, 2001).

Disregarding the ASQ, cases with more than 30% missing data on the scales were initially excluded from the analyses. For the student sample 5 cases were excluded, while in the general population sample 4 cases were excluded. The total number of subjects reported for the two samples (N=875 and N=303, respectively) is the total number of cases included in the analysis after exclusion of cases with missing data. Remaining missing values were replaced by series mean. As mentioned, the total N for the ASQ was 249 due to shortage of time for 54 of the students.

The understanding of norm applied in this study is norm as population average (Kendall & Sheldrick, 2000). Thus, in line with a dimensional and continuous understanding of depression we will not apply any score-based exclusion criteria on the data in order to create what Kendall & Sheldrick (2000) refer to as a 'supernormal' reference group without cases with elevated symptomatology.

Results

Descriptive and correlational data for the general population sample

From table 5a it can be seen that the scale mean for the BDI-II was 8.12 (SD=7.5) for the general population sample. This mean is outside the symptomatic range recommended by Beck et al. (1996). Table 6 presents total BDI-II raw scores converted to z-scores.

Insert table 5a and b and table 6 about here

Examination of the item means reveals that item means range from .08 for item 9 (suicidal thoughts) to .75 for item 15 (loss of energy). In addition to item 9, item 6 (punishment feelings) and item 1 (sadness) have low means, while items 16 (sleep changes) and item 20 (fatigue) have the highest means in addition to item 15 (loss of energy). The corrected item-total correlations showed that each individual item contributed substantially to the total BDI-II score (>.35). The item-total correlations were ranging from

.38 for item 21 (loss of sexual interest) and 6 (punishment feelings) to .66 for item 7 (self-dislike) for the general population sample. The mean inter-item correlation was .326 (table 5b).

A 2x4 analysis of variance of mean total BDI-II score by age group and sex in the general population sample (table 7) showed a significant effect of sex ($F(1, 525)=9.5$, $p<.002$). There was no age effect, ($F(3, 143)=2.6$, $p>.051$, age was also correlated with the total BDI-II score resulting in a nonsignificant result ($r=-.026$, $p>.440$.)), nor interaction effect ($F(3, 66)=1.2$, $p>.30$).

Insert table 7 about here

Figure 1 shows the general population sample and student sample divided into symptom severity groups based on Beck et al. (1996) recommendation for interpretation of the total BDI-II score. According to the results, 81.0 % of the general population sample respondents reported minimal symptomatology, 10.6 % reported mild symptomatology, 6 % reported moderate symptomatology and 2.4 % of the respondents reported severe symptomatology.

Insert figure 1 and table 9 about here

Descriptive and correlational data from the student sample

The total BDI-II scale mean for the student sample was 7.12 (SD=6.0)(Table 8a), which was not significantly different from the scale mean of the general population sample ($t(1, 1176)=1.12$, $p=.264$). Item means inspection reveals that items 16 (sleep changes), 5 (guilt feelings) and 15 (loss of energy) have the highest, while 9 (suicidal thoughts), 1 (sadness), 14 (worthlessness) and 21 (loss of interest in sex) have lowest means. Item-total correlations for the student sample were ranging from .31 item 17 (irritability) to .61 item 10 (crying). Four items (6, 8, 17 and 21) had moderate item-total correlations (<.35). The mean inter-item correlation was .337 (table 8b).

Insert table 8 a and b about here

As can be seen from table 10, the mean total BDI-II score was higher for women than for men (7.4 vs. 6.3), but contrary to the result from the general population sample, this difference was not statistically significant (t(1, 301)=-1.4, p>.155).

Insert table 10 and 11 about here

Table 11 and figure 1 illustrate the student sample separated into groups based on symptom severity. Fewer respondents reported symptoms of depression in the student sample than in the general population sample. Altogether 89 % reported minimal symptomatology, 5.8 % report mild symptomatology, 4.2 % reported moderate symptomatology and 1.0 % reported severe symptomatology.

Reliability analyses

The internal consistency estimates for the BDI-II were high in both samples in the present study. The corrected alpha value was .91 for the general population sample and .86 for the student sample (table 5a and 8a). Further analysis of the reliability of the scale was provided by the three-week test-retest correlation for the BDI-II in the student sample, which was .77 (p<.001).

Convergent and discriminant validity analyses

In table 11 bivariate correlations (Pearson's r) between the BDI-II and the convergent and discriminant scales are listed. The correlational analysis shows that the BDI-II correlates significantly with the Zung-SDS (r=.71, p<.001) and the SCL-90 depression subscale (r=.84, p<.001).

Insert table 12 about here

The analysis further revealed that the BDI-II correlates significantly with the BAI (r=.67, p<.001) in the student sample, and the general population sample (r=.62, p<.001). Additionally, the correlation between the BDI-II and the SCL-90 psychotisism subscale was lower but significant (.61, p<.001).

Subsequent analysis of the differences between the correlations showed that the correlation between the BDI-II and the convergent measures (r=.71, Zung-SDS and r=.84, SCL-90 dep.) were significantly higher than the correlation with discriminant measure of SCL-90 psychotisism scale (r=.61)(p<.028 and p<.001, respectively). The correlation between the BDI-II and the BAI (r=.67), however, were not significantly lower than the correlation with the Zung-SDS (p>.346).

Construct validity analyses

The mean scores for the ASQ scale are presented in table 13. As can be seen by the pattern of the mean ASQ scores, all respondents showed a tendency toward internal, stable and global attributions for both positive and negative events. For the positive events the mean scores were 4.7 (SD=0.6), and 5.0 (SD=0.7) and 4.7 (SD=0.8) for internality, stability and globality respectively. The corresponding means for negative events were 4.2 (SD=0.7), 4.2 (SD=0.9) and 3.9 (SD=1.0). Thus, the only exception from the pattern was local vs. global attributions for negative events (mean=3.9), which was going slightly towards local attributions. The mean score of the importance scale revealed that the respondents perceived both positive and negative events as important to them, with the positive events rated as the most important (mean scores of 5.5 (SD=0.8) and 4.7 (SD=1.0), respectively).

Insert table 13 about here

The respondents who completed the ASQ and the BDI-II (N=249) were then separated into two groups based on the total BDI-II score (total BDI-II score =<13). The mean scores for attributions by BDI-II total score group is presented in table 14. Inspection of the means show that, in line with the expectations, the respondents with elevated symptomatology produced higher mean scores than the unsymptomatic group on all attributions for negative events, and lower mean scores attributions for positive events. A subsequent analysis of variance, however, revealed that there were significant group differences only for internal and stable attributions for positive events (F(1, 247)= 9.3, p<.002, and F(1, 247)= 27.4, p<.001, respectively) and internal and global attributions for negative events (F(1, 247)= 6.7, p<.01, F(1, 247)=4.2, p<.04, respectively).

Insert table 14 about here

Confirmational factor analyses

Adopting West, Finch & Curran (1995, cited in Osman et al. 1998) suggestion for acceptable ranges for skewness (<2) and kurtosis (<4), five of the items in the general population sample had higher skewness and kurtosis than the suggested ranges (table 5a), and in the student sample seven of the items had too high skewness and six had too high kurtosis (table 8a). In order to avoid violation of the assumption of normality in Structural Equation Modelling, a square root transformation of the data was performed on both sample data before the factor analysis was carried out (Ferguson & Takane, 1989). After the transformation the data produced acceptable skewness (<2) and kurtosis (<4) in both samples.

The fit of the following models were then evaluated in both samples employing a maximum-likelihood estimation procedure in Amos 4:

1.*a two-factor model* based on the oblique two-factor structure reported by Beck et al. (1996), Dozois et al. (1998) and Whisman (2000). A cognitive-affective factor and a somatic factor were specified as latent variables, and 16 items were constrained to load on the cognitive-affective factor and 5 items constrained to load on the somatic factor. The factors were allowed to correlate.

2.*a three-factor model* based on Osman et al.'s (1998) modification of Byrne & Baron (1993) analyses of the original BDI. The factor names were 'negative attitude', 'performance difficulties' and 'somatic'. In line with Osman et al. (1998), and different from model 1., item 11, 16 18 and 21 were constrained to load on the somatic factor in this model. The factors were allowed to correlate.

3.*a second-order model* with a general second-order factor 'depression' and three first-order factors similar to those hypothesized in model 2. The covariation among the three first-order factors was hypothesized to be explained fully by their regression onto the second order factor.

Insert figures 2-4 about here

The hypothesized models are presented graphically in figures 2-4. The fit indices of the models tested on the general population data are presented in table 15. Disregarding the chi-square due to sample size, the fit indices revealed that all the three hypothesized models produced good fit to the general population sample data (CFI=.91 and RMSEA<.10). Based on the AIC index, the three-factor model was then retained as the model that parsimoniously best fit the data (AIC=768.4).

Insert tables 15 and 16 about here

In order to improve the fit of the three-factor model, the model was modified in a post-hoc analysis (Byrne, 2001). Following theoretically meaningful modifications were carried out: error variances between pairs of item 1(sadness) and 10 (crying), item 4 (loss of pleasure) and 12 (loss of interest), and item 15 (loss of energy) and 20 (fatigue) were allowed to correlate (Osman et al. (1998), Whisman et al. (2000)). The fit indices of this modified three-factor model for the general population sample were all indicating excellent fit: CFI=.95 and RMSEA=.046 (table 16). The model is presented graphically in figure 5 with the belonging standardized regression weights. As can be seen, item loadings were ranging from .44 (item 6) to .76 (item 7). The correlation between the 'negative attitude' and 'somatic' factors was .81, indicating a 65% overlap of the variance in the two latent variables. The

correlation between 'negative attitude' and 'performance difficulties' were .83 (69% overlap), and the correlation between 'performance difficulties and 'somatic' was as high as .99 (98% overlap). The alpha values for the factors were .86 for 'negative attitude', .84 for 'performance difficulties' and .58 for 'somatic'.

Insert figure 5 about here

The modified three-factor model was then tested on subgroups of the general population sample in order to test the robustness of the factor model across different population groups. Table 16 summarizes the results for the general population divided into subgroups for sex, education and marital status. As indicated by the fit indices, the model produced satisfactory fit for all groups (CFI>.90, RMSEA<.10).

Insert table 17 and figure 6 about here

Table 17 summarizes the fit indices for the student sample. As was the case for the general population sample, all the three models produced very similar fit indices. However, all the CFI indices of the three models are below the pre-established criteria of .90. Based on the AIC index, the three-factor model was again retained and modified in order to achieve a better fitting model (AIC=444.3). Following modifications were carried out: first, the same pairs of items' error variances were allowed to correlate as in the modified model for the general population sample. Additionally, item 7 (self-dislike) and 8 (self-criticalness) were allowed to correlate (Whisman et al., 2000) and item 21 (loss of sexual interest) were constrained to load on the 'performance difficulties' factor (!). The modified model is presented in figure 6, and as can be seen in table 17, the model produced satisfactory fit indices (CFI=.90 and RMSEA=.048). The regression weights were ranging from .39 for item 21 (loss of sexual interest) to .64 for item 2 (pessimism). The correlations between the factors were .52 (negative attitude x somatic), .72 (negative attitude x performance difficulties) and

.69 (performance difficulties x somatic). The alpha values for the factors were .83 for 'negative attitude', .78 for 'performance difficulties' and .56 for the 'somatic' factor.

Discussion

Normative data for the Norwegian BDI-II

A central aim of this investigation is to contribute to the development of assessment methods for use in Norway by providing normative data for a Norwegian version of the BDI-II. Additionally, no other studies have employed the BDI-II with a general population sample, and Beck et al. (1996), and Steer & Clark (1997) have recommended that the normative basis of the BDI-II must be investigated in a variety of non-clinical samples, not just in student samples.

The reported scale mean for the Norwegian general population sample was lower but comparable to reported means from the earlier studies on the BDI-II with student samples. The closest comparable means are reported in the studies by Whisman et al. (2000) and O'hara et al. (1998). The mean values for individual items differed to some extent from the item means reported on the student sample in the manual (Beck et al., 1996). The exceptions are item 9 (suicidal thoughts) and item 6 (punishment feelings), which were amongst the lowest also in Beck et al's (1996) sample, and item 16 (sleep changes), 20 (fatigue) and item 15 (loss of energy), which were amongst the highest. These findings seem meaningful considering the more pathological content of the two first items, and the less pathological nature of the latter.

Further in line with Beck et al. (1996) and Osman et al. (1998), but contrary to three of the other reviewed studies (Steer & Clark, 1997, O'hara et al., 1998, Whisman et al., 2000), we found a significant sex difference with women reporting higher scale mean than men in the general population sample.

Adding to the normative data for student populations, the scale mean for the student sample was not signifiantly different from the mean in the general population sample and albeit comparable to earlier findings with student samples. No sex differences were found among the students, a result which confirms earlier research reporting that sex differences are

more seldom in student samples than in the general population (Nolan & Wilson, 1994). One sample specific item mean was noteworthy: the mean of item 21 (loss of sexual interest) is much lower for the student population than the general population sample.

Sample considerations

The normative data of the BDI-II scale obtained in the general population sample may be influenced by the sampling procedure applied. As explained in the methods section, participants in the general population sample were recruited through private telephone numbers and a subsequent letter that had to be returned with the completed questionnaires. Although the response rate obatined in the present study is comparable to other general popualtion studies applying this procedure (Cohen et al., 1996), the representativeness of the sample in the present investigation can be questioned. One important factor influencing the sampling is that no inpatients are included in the sample with the applied procedures. Thus, individuals with seriously elevated symptomatology are not represented in the data. Taking further into account the functional impairment aspects of depression, another related effect that may have influenced our sampling could be that individuals with elevated depressive symptomatology may not respond to the phonecall. A similar effect may emerge even though they respond. After being briefed about the survey, respondents may refuse to participate due to depression symptoms like fatigue and loss of energy. As a result, respondents with less serious, but elevated, depressive symptomatology may therefore be under-represented in the general population sample. Accordingly, the scale mean may be lower for our sample than what is truly representative for the general population.

Another consideration relevant to the scale means of both samples is on the positively skewed distribution of the scores. Although the skewness obtained in the present investigation is comparable to earlier research with non-clinical samples (Osman et al., 1998, Dozois et al., 1998, Whisman et al., 2000), several authors (Joiner et al., 1994, Kendall & Sheldrick, 2000) have questioned whether low-end scores on the BDI actually reflect absence of depression symptomatology. As mentioned in the introduction, a disadvantage with self-report scales like the BDI-II is that they are relatively face valid, which makes them easier to fake. For example, Joiner et al. (1994) have showed in a study that low-end scorers on the BDI-II have elevated scores on the MMPI-L scale, suggesting a tendency to deny

symptoms. This tendency may have again caused a lower scale mean for our samples than what is really true.

However, the tendency to deny symptoms should be especially potent if there are personal consequences involved in filling out the self-report (Joiner et al., 2000). This was not the case for neither of the samples in the present study, and it is therefore less likely that this effect may have had large impact on the results.

Psychometric characteristics of the BDI-II

Reliability. In line with earlier research on the BDI, and previous studies on the BDI-II, the reliability coefficient (alpha value) was expected to be high for both samples. In line with this expectation, alpha values of .91 and .86 were estimated for the general population sample and student sample respectively. Both values are comparable to earlier findings with non-clinical samples and can be considered as evidence for high internal consistency of the Norwegian BDI-II.

Boyle (1985) has argued earlier, however, that there is a misunderstanding in the literature on self-report scales of depression that the internal consistency should be as high as possible. Boyle (1985) states that if the internal consistency of the scale is too high, every new item does not add much new information. Along the same lines he continues by arguing that most self-report scales of depression are not broad measures of depression, considering their consistent high internal consistency. Thus, taking into account this perspective, it can be argued that the high internal consistency of the BDI-II found in this study supports the validity of the scale, but may also suggest that the scale is a rather narrow measure of depression.

No studies to date have investigated the test-retest of the BDI-II with a student sample. The three week test-retest correlation in the present study was comparable to the three week test-retest correlation reported by Oliver & Burkham (1979) on the original BDI, and lower than the one-week test-retest correlation of reported by Beck et al. (1996) with their small outpatient sample.

According to Kline (2000) a test-retest should be at least .80 over a three-month period in order to reflect high reliability, and it can therefore be argued that the three week test-retest reliability found in the present study is too low. On the other hand, the test-retest correlation of a measure should also be evaluated with respect to whether it is supposed to be sensitive to

the present state or tap a more trait-like construct. As far as the BDI-II is concerned, the scale is essentially a state measure and should therefore be sensitive enough to detect differences in depression under different circumstances and over a time-span extending the two weeks specified in the instructions. This latter characteristic is especially important regarding that the BDI-II is often used as a therapy outcome evaluation measure that should be sensitive to symptom changes over relatively short time periods. Thus, a test-retest of .77 obtained in the present study may indicate that the BDI-II is a sufficiently reliable measure, but also appropriately sensitive to actual changes in depression symptoms in non-clinical subjects over a three-week period.

Convergent and discriminant validity. In order to investigate the convergent and discriminant validity of the BDI-II, the SCL-90-R depression subscale and the Zung-SDS were employed as convergent measures, and the BAI and the SCL-90-R psychotisism subscale were employed as discriminant measures in the present investigation. The correlational analyses showed that, in line with our expectations, the BDI-II correlated highly and significantly with the SCL-90-R depression subscale and the Zung-SDS. The correlation between the BDI-II and the SCL-90-depression subscale is higher than earlier reported correlation by Rouseville (1979, cited in Beck et al., 1988), while the correlation with Zung-SDS is exactly the same as the mean correlation reported from Beck et al. (1988) review on the original BDI.

The correlation between the BDI-II and the SCL-90-R psychotisism subscale was comparable to Gotlib (1984) finding with the original BDI. As expected, this correlation was significantly lower than the correlations between the BDI-II and the two convergent measures, thus indicating convergent and discriminant validity of the BDI-II with these measures respectively.

The BAI has often been employed as a discriminant measure in relation to the BDI (Beck et al., 1988), but there is some disagreement in the literature related to the discriminant power of the inventory in relation to the BAI. The correlations between the BDI-II and the BAI found in the present study were higher, but comparable, to the correlations found earlier (Beck et al., 1996, Steer & Clark, 1997). Subsequent analysis showed that the correlation between the BDI-II and the BAI in the student sample was not significantly lower than the correlation between the BDI-II and the Zung-SDS, a result which suggest that the BAI should perhaps be regarded as a convergent instead of a discriminant measure (Steer & Clark, 1998).

On the other hand, it may be argued that all the correlations found among the self-report measures in this study is rather high. Overall, none of the correlations between the BDI-II and the other measures were below .61, which indicate that self-report scales may not discriminate very well among each other. Theoretically, advocates of a categorical perspective on depression (Coyne, 1994, Gotlib, 1984 and Gotlib & Hammen, 1992) would argue that these results reflect neither convergent nor discriminant validity, but are more in line with the hypotheses of general distress (Gotlib, 1984) or negative affectivity (Watson & Clark, 1984) in non-clinical subjects. Adopting a similar line of reasoning, it is possible that the correlations between the BDI-II and the other measures in the student sample imply that the BDI-II has low specificity for depression in non-clinical samples.

Construct validity. There is some support for the construct validity of the BDI-II (Osman et al., 1998), but no studies have employed the BDI-II together with the ASQ. In line with the reformulated hopelessness theory of depression, we expected BDI-II scores above 13 to be associated with increasing internal, stable and global attributions for negative events and the opposite pattern for positive events. The analyses of the attribution dimensions revealed that for positive events, attributions of internality and stability discriminated between the BDI-II groups, and for negative events attributions of internality and globality discriminated between the BDI-II groups. Thus, partly in line with our hypotheses, the results suggest evidence for attribution of internality as associated with depression symptoms above 13, but only limited evidence for stability and globality attributions.

A factor that further questions the findings is the low alpha values for the internality attribution scale. Because of the low internal consistency of this subscale, an attribution index based on all attribution dimensions should perhaps have been computed and included in the analysis in order to create a more reliable scale. On the other hand, the interpretation of such a composite scale would not been clear since the three attributional dimensions may not have equivalent status. Supporting this reasoning, Abramson, Alloy & Metalsky (1989) have suggested that stability and globality attributions may have a more direct relationship with depression than internality.

With respect to the low reliability of the internality dimension it can be argued that the limited evidence for construct validity found in the present investigation is due to the scale characteristics of the Norwegian version of the ASQ, and not the BDI-II. Although used with several clinical samples (Hoffart & Martinsen, 1991, Hoffart & Torgersen, 1991) the Norwegian version of ASQ has not been translated and validated for use in Norway

according to recommended procedures (Geisinger, 1994, Cohen et al., 1996). This should be carried out before the ASQ is applied as a measure of attribution style in clinical and non-clinical samples.

Factor structure of the BDI-II

The present study aimed to investigate the factor structure of the Norwegian BDI-II in a general population sample and thereby extend the work on the factor structure of the BDI-II. Based on the work of Beck et al. (1996), Dozois et al. (1998), Osman et al. (1998) and Baron & Byrne (1993), confirmational factor analyses were carried and three hypothesized models were tested on both samples. For the general population sample all the three models produced acceptable fit indices, indicating that several models fit the data from the BDI-II adequately. Equal fit of several models as obtained in the present study is often found in CFA (MacCallum, Wegener, Uchino & Fabrigar, 1993). Taking into account the parsimony of the models, however, the three-factor model was retained for post-hoc analyses. This three-factor model of the BDI-II is equivalent to the three-factor model reported by Osman et al. (1998), thus establishing further support for a 'negative attitude' factor, a 'performance difficulties' factor and a 'somatic' factor parsimoniously explaining the results from the BDI-II with non-clinical samples.

Several modifications were needed in order to obtain a model with robust fit across the subgroups of the general population sample. After estimating three pairs of theoretically meaningful correlated residuals, the three-factor model produced significant fit for population subgroups by sex, education and marital status, indicating a robust and generalizable three-factor model of the BDI-II.

The three-factor model was also retained for the student sample. However, additional interesting modifications were needed in order to obtain a satisfactory fit of the three-factor model to the student data. Only when item 21 (loss of sexual interest) was constrained to load on 'performance difficulties' factor instead of the 'somatic' factor, did the three-factor model fit the data satisfactory.

As mentioned in the introduction, Beck et al. (1996) reported a low contribution of item 21 (loss of sexual interest) to the total scale score, and a nonsignificant factor loading in the exploratory factor analyses. Dozois et al. (1998) and Whisman et al. (2000), however, found that this item loaded on the cognitive-affective factor. In line with Osman et al. (1998), item

21 (loss of sexual interest) was initially constrained load on the somatic factor in the present study, but subsequently modified in the post-hoc analyses to load on the 'performance difficulties' factor in the student sample. The item also produced among the lowest standardized regression weights in the factor analyses for both samples. Seen as a whole, the low regression weights obtained for this item, and the modification analyses of the three-factor model on the student data, are to some extent contradicting the findings of Dozois et al. (1998) and Whisman et al. (2000), but confirming the somewhat unclear status of this item found by Beck et al. (1996) in the manual.

A possible explanation for the modification of item 21(loss of sexual interest) to load on the 'performance difficulties' factor with the student sample may be that the mean age of this sample is much lower than for the general population sample. It may be that for younger people loss of interest in sex is not so much a question of a reduced somatic ability, as it is a result of a more motivational or psychological factor involved in depression. This explanation is preliminary, however, since the factor structure for the BDI-II was not tested on age groups in the general population sample. Further research in different samples and different age groups is certainly needed in order to clarify the status of item 21(loss of sexual interest) in the BDI-II.

The high item-total correlations and regression weights for item 7 (self-dislike) is confirming earlier findings by Beck et al.(1996), Dozois et al.(1998) and Whisman et al. (2000). This consistent finding is noteworthy considering that this symptom is not included in the DSM-IV criteria for Major Depressive Episode. As mentioned in the introduction, Beck and colleagues (Clark & Beck, 1999) have critized the DSM-IV manual for the strong emphasis on somatic symptoms to the relative neglect of subjective symptoms. If the consistent finding that the subjective item 'self-dislike' is highly correlated with both the total score and the latent cognitive factor can be replicated with clinical samples, this supports Clark & Becks (1999) argument that the weighting of subjective versus somatic criteria in the manual is not supported by empirical evidence, and further indicate that this symptom should be included in themanual.

A final concern relates to the high correlation between the 'performance difficulties' factor and the 'somatic' factor found in the general population sample, and the low reliability for the somatic factor found in both samples. The high correlation between the factors suggests that the overlap between these two latent variables is high, which again indicate that the items of these two factors can be merged into one factor. Theoretically, a joint factor

consisting of the items loading onto the 'performance difficulties' and 'somatic' factors can be interpreted as a somatic/motivational dimension in depression. Psychometrically, this factor solution would be more parsimonious than the original three-factor structure. Furthermore, construction of a new factor would also solve the problem with the low reliability of the somatic factor, since this new factor is likely to have much higher reliability. No studies to date have investigated this hypothesized factor structure, and future investigations of the factor structure of the BDI-II should be attentive to this possible factor structure.

Summary and conclusions

This investigation aimed to provide normative data and investigate the psychometric characteristics and factor structure of a Norwegian version of the BDI-II. As far as normative data are concerned, it seems reasonable to conclude that paying appropriate attention to the methodological issues discussed concerning sampling procedure and symptom denial, the results from the present investigation provide useful normative comparison data for the BDI-II from a sample of the Norwegian population and a student sample. The normative data are further supported by the high internal consistency of the BDI-II in both samples and sufficient test-retest reliability over a three-week period in the student sample. In addition, the convergent and discriminant validity results are comparable to earlier research, but the interpretation of these results is dependant on the perspective on depression. Adopting a dimensional perspective (Clark & Beck, 1999) the results can be interpreted as evidence for convergent and discriminant validity with depression and psychotisism scales respectively, but are less conclusive as support for the discriminative validity with the anxiety scale. The construct validity results are promising, but needs to be investigated further with a measure standardized for use in Norway. Finally, strong evidence for a modified three-factor model of the BDI-II was obtained across different subsamples of the general population.

Seen as a whole, the results of these analyses provide considerable empirical support for a Norwegian version of the BDI-II with non-clinical samples. The Norwegian BDI-II should now be further examined with respect to normative data, psychometric characteristics and factor structure with clinical samples.

Beyond the validation of the Norwegian BDI-II, this investigation was designed to extend research on the BDI-II in general, with special attention to the factor structure of the scale. The majority of research on the BDI-II is conducted with student samples and the

present investigation extends research on the BDI-II by providing normative data and psychometric characteristics from a general population sample. Moreover, no other studies have investigated the test-retest reliability of the BDI-II with non-clinical samples, and the present study provides sufficient support for the test-retest reliability of the BDI-II with a student sample. Convergent and discriminant validity of the BDI-II is also established with measures never before applied with the BDI-II, but more research on the continuity of depression symptoms as assumed within a dimensional perspective on depression in needed in order to make definite conclusions on this issue. Research on both clinical and non-clinical samples may provide more insight into this contentious issue (Flett et al., 1997).

As mentioned, very few studies have employed CFA on the BDI-II and no studies have employed CFA on the BDI-II with a general population sample. The present investigation therefore extends research on the BDI-II in general by providing further support for a three-factor model of the BDI-II. The three-factor model retained is in line with the results from the study by Osman et al. (1998), but additional modifications were needed in to obtain satisfactory fit in the subgroups of the general population sample and the student sample. It is also important to note that since all the three models tested produced satisfactory fit in the general population sample, the results do not disconfirm the two-factor solution found by Beck et al. (1996), Steer & Clark (1997), Dozois et al. (1998), nor the second-order structure based on the work of Baron & Byrne (1993). Rather, the results in the present investigation indicate that several models fit the BDI-II data from a general population sample, but a three-factor model was retained as the model that parsimoniously best explained the data.

Further research on the factor structure of the BDI-II should be carried out in both non-clinical and clinical samples in order to determine the generalizability of the three-factor solution obtained in this study. Attention should be given the status of item 21 (loss of sexual interest), especially for different age groups, and item 7 (self-dislike) in clinical samples. Finally, a possible alternative two-factor solution of the BDI-II suggested in the present study may be investigated further.

References

Abramson, L., Alloy, G.I., Metalsky, G.I. (1989). Hopelessness depression – a theory-based subtype of depression. *Psychological review, 96*, 358-372.

Abramson, L., Seligman, M.E.P. & Teasdale, J.D. (1978). Learned Helplessness in Humans: Critique and Reformulation. *Journal of Abnormal Psychology, 87*, 49-74.

Akaike, H. (1987). Factor analysis and AIC. *Psykometrik*a, 52, 317-332.

American Psychiatric Association (1994). Diagnostic and Statistical Manual of mental disorders (4th edition). Washington, DC; APA.

Arbuckle, J.L. (1995). Amos 4. User's Guide. Smallwaters Corporation.

Barret, J.E., Barret, J.A., Oxman, T.E., Gerber, P.D. (1988). The prevalence of psychiatric disorders in primary care practice. *Archives of General Psychiatry, 45*, 1100-1106.

Bech, P. (1993). Rating scales for Psychopathology, Health status and Quality of Life. A Compendium on Documentation in Accordance with the DSM-III-R and WHO systems. Springer-Verlag; 1993.

Beckham, E.E. & Leber, W.R. (1995). Handbook of depression. Second Edition. The Guilford Press, New York.

Beck, A.T (1967). Depression; Clinical, Experimental and Theoretical Aspects. University of Pennsylvania Press, 1967.

Beck, A.T. (1982). Situational Anxiety Check List. University of Pennsylvania, Center for Cognitive Therapy, Philadelphia.

Beck, A.T., Rush, A.J., Shaw, B.F. & Emery, G. (1979). Cognitive Therapy of Depression. New York: Guilford Press.

Beck, A.T. & Steer, R.A. (1993a). Manual for the Beck Anxiety Inventory. San Antonio, TX: Psychological Corporation.

Beck, A.T., Steer, R.A. & Brown, G.K. (1996) Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation.

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

Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Boyd, J.H., Weissman, M.M., Thompson, W.D. & Myers, J.K. (1982). Screening for depression in a community sample: Undersatnding the discrepancies between depression syndrome and diagnostic scales. *Archives of General Psychiatry, 39* (10), 1195-1200.

Boyle, G.J. (1985). Self-report measures of depression: some psychometric considerations. *British Journal of Clinical Psychology, 24*, 45-59.

Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K.A. & Long, J.S. (Eds.) Testing structural equation models. Newbury Park, California: Sage, 136-162.

Buchwald, A.M. & Rudick-Davis, D. (1993). The symptoms of major depression. *Journal of Abnormal Psychology, 102*, 197-205.

Byrne B.M. & Baron, P. (1993). The Beck Depression Inventory: Testing and cross-validating an hierarchical factor structure for nonclinical adolescents. *Journal of Counselling and Development, 26*, 164-178.

Byrne, B.M., Baron, P., Larsson, B. & Mehlin, L. (1995). The Beck Depression Inventory: Testing and cross-validating a second-order factorial structure for Swedish nonclinical adolescents. *Behviour Research and Therapy, 33*, 345-356.

Byrne, B.M. (2001). Structural Equation Modelling with Amos. Basic Concepts, Applications and Programming. Multivariate Applications Book series. Lawrence Erlbaum, New Jersey.

Cicchetti, D.V. (1994). Guidelines, Crietria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychological Assessment, 6*, 284-290.

Clark, D.A. & Beck, A.T. (1999). Scientific Foundations of Cognitive Theory of Depression. John Wiley & sons. New York, 1999.

Clark, D.A., Watson, R. & Reynolds, J.C. (1995). Diagnosis and classification of psychopathology – challenge to the current system and future directions. *Annual Review of Psychology, 46*, 121-153.

Clark, L.A. & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxomonic implications. *Journal of Abnormal Psychology, 100*, 316-336.

Cohen, J.C., Swerdlik, M.E. & Phillips, S.M. (1996). Psychological Testing and Assessment. An Introduction to Tests and Measurement, Third edition. London, 1996: Mayfield.

Costello, C.G. (1993). Symptoms of Depression. John Wiley & Sons. New York.

Coyne, J.C. (1994) Self-reported distress: Analog or ersatz depression? *Psychological Bulletin, 116*, 29-45.

Derogatis, L.R (1983). SCL-90-R. Administration, scoring and procedures manual. Baltimore, Clinical Psychometric Research.

Derogatis, L.R. & Dellapietra, L. (1994). Psychological Tests in Screening for Psychiatric Disorder. In Maruish (1994): <u>The use of Psychological Testing for Treatment Planning and Outcome Assessment</u>. Lawrence Erlbaum; New Jersey: 1994.

Dozois, D.J.A., Dobson, K.S. & Ahnberg, J.L. (1998) A psychometric evaluation of the Beck Depression Inventnory-II. *Psychological Assessment, 10*, 83-89.

Ennis, J., Barnes, R.A., Kennedy, S. & Trachtenberg, D.D. (1998). Depression in self-harm patients. *British Journal of Psychiatry, 154*, 41-47.

Enns, M.W., Cox, B.J. & Borger, S.C. (2001). Correlates of analogue and clinical depression: a further test of the phenomenological continuity hypothesis. <u>Journal of Affective Disorders, 66</u>, 175-183.

Farmer, G. & Nelson-Gray, T. (1990). The accuracy of counting vs. estimating event frequency in behavioural assessment: The effects of behavioural frequency, number of behaviours monitored and time delay. *Clinical psychological review, 10*, 453-460.

Fechner-Bates, S., Coyne, J.C. & Schwenk, T.C. (1994). The relation of self-reported distress to depression disorder and other psychopathology. *Journal of Counselling and Clinical psychology, 62*, 550-564.

Flett, G.L., Vredenburg, K. & Krames, L. (1997). The Continuity of Depression in Clinical and Non-clinical Samples. *Psychological Bulletin, 121* (3), 395-416.

Geisinger, K.F. (1994). Cross-Cultural Normative Assessment: Translation and Adaption Issues Influencing the Normative Interpretation of Assessment Instruments. *Psychological Assessment, 6*, 304-312.

Gotlib, I.H. (1984). Depression and General Psychopathology in University Students. *Journal of Abnormal Psychology, 93*, 19-30.

Gotlib, I.H. & Hammen, C.L. (1992). <u>Psychological Aspects of Depression. Toward a Cognitive- Interpersonal Integration</u>. New York: John Wiley & Sons, 1992.

Hamilton, M. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology, 32*, 50-55.

Hoffart, A. (1985). <u>Norsk versjon av Attributional Style Questionnaire</u>. Modum Bads Nervesanatorium.

Hoffart, A. & Martinsen, E.W. (1991). Causal Attributions in Clinical Subtypes of Depression: A Longitudinal Study of Inpatients. *Journal of Psychopathology and Behavioural Assessment, 13*, 241-248.

Hoffart, A. & Torgersen, S. (1991). Causal Attributions in First-Degree Relatives of Depressed and Agoraphobic Inpatients. *Comprehensive Psychiatry, 32*, 458-464.

Joiner, T.E. (1996). A confirmatory Factor-Analytic Investigation of the Tripartite Model of Depression and Anxiety in College Students. *Cognitive Therapy and Research, 20*, 521-539.

Joiner, T.E., Schmidt, K.L. & Metalsky, G.I. (1994). Low-end specificity of the Beck Depression Inventory. *Cognitive Therapy and Research, 18*, 55-68.

Joiner, T.E., Schmidt, N.B., Cook, J.H, Gencoz, F. (2000). Differential Roles of Depressive and Anxious Symptoms and Gender in defensiveness. *Journal of Personlity Assessment, 75* (2), 200-211.

Katz, R., Shaw, F.B., Vallis, T.M. & Kasier, A.S. (1995). The Assessment of Severity and Symptom Patterns in Depression. In Beckham & Leber (eds.). Handbook of Depression. Second Edition. The Guilford Press, New York.

Jøreskog, K.G. (1993). Testing structural equation models. In K.A. Bollen & J.S. Long (Eds.) Testing structural equation models. Sage Publications.

Katz, R., Shaw, T. Vallis, M. & Kaiser, A.S. (1995). The Assessemnt of Severity and Symptom Patterns in Depression. In Beckham, E.E. & Leber, W.R. (Eds.) Handbook of Depression. Second Edition. The Guilford Press, New York.

Kendall, P.C. & Sheldrick, R.S. (2000). Normative data for normative comparisons. *Journal of Counselling and Consulting Psychology, 68*, 767-773.

Kessler, R.C., McGonagh, K.A., Zhao, S.Y., Nelson, C.B., Hughes, M., Eshleman, S., Wittchen, H.U. & Kendler, K.S. (1994). Lifetime and 12-month prevalence of DSM-III pscychiatric disorders in the US: Results from the National Comorbidity Study. *Archives of General Psychiatry, 51*, 8-19.

Kline, P. (2000). The Handbook of Psychological Testing. Second Edition. Routledge, London 2000.

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.

MacCallum, R.C., Wegener, D.T., Uchino, B.N., Fabrigar, L.R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114*, 185-199.

Maruish, M.E. (1994). The use of psychological testing for treatment planning and outcome assessement. New Jersey, 1994: Lawrence Erlbaum Associates.

Maruish, M.E. (2000). Handbook of Psychological Assessment in Primary Care Settings. Lawrence Earlbaum; New Jersey.

Maser, J.D, & Cloninger, C.R. (1990). Comorbidity of Mood and Anxiety Disorders. Washington DC, American Psychiatric Press, 1990.

Moran, P.W. & Lambert, M.J. (1983). A review of current assessment tools for monitoring changes in depression. In M.J. Lambert, E.R. Christensen & S.S. DeJulio (Eds.) The assessment of psychotherapy outcome (304-355). New York; John Wiley.

Myers, J.K. & Weissman, M.M. (1980). Use of a self-report symptom scale to detect depression in a community sample. *American Journal of Psychiatry, 137* (9), 1081-1084.

Nezu, A.M., Ronan, G.F, Meadows, E.A. & McClure, K.S. (2000). Practitioner's guide to Empirically Based Measures of Depression. Kluwer Academic, New York, 2000.

Nolan, R. & Wilson, V.L. (1994). Gender and Depression in an undergraduate population. *Psychological Reports, 75*, 1327-1330.

Nordhagen, T. (2001). Hovedoppgave i psykologi, Psykologisk fakultet, Universitetet i Bergen, 2001.

Nordhus, I.H., Pallesen, S. & Aasen, H. (2001). Norsk versjon av Beck Depression Inventory-II. Institutt for Klinisk Psykologi, Psykologisk fakultet, Universitetet i Bergen.

Nordhus, I.H., Pallesen, S. & Nordhagen, T. (2001). Norsk versjon av Beck Anxiety Inventory. Institutt for Klinisk Psykologi, Psykologisk fakultet, Universitetet i Bergen.

Ayuso-Mateos, J.L., Vazquez-Barquero, J.L., Dowrick, C., Lehtinen, V., Dalgaard, O.S., Casey, P., Wilkinson, C. Lasa, L., Page, H., Dunn, G. & Wilkinson, G. (2001) Depressive disorders in Europe: prevalence figures from the ODIN study. *British Journal of Psychiatry, 179*, 308-316.

O'hara, M.M., Sprinkle, S.D. & Ricci, N. (1998). Beck Depression Inventory-II: College population study. *Psychological Reports, 82*, 1395-1401.

Opinion. (2001). Rapport fra landsomfattende omnibus, oktober og november, 2000.

Osman, A., Downs, W.R., Barrios, F.X., Kopper, B.A., Gutierrez, P.M. & Chiros, C.E. (1997). Factor structure and psychomtreic characteristics of the Beck Depression Inventory-II. *Journal of Psychopathology and Behavioural Assessment, 19*, 359-376.

Peterson, C., Semmel, A. von Baeyer, C., Abramson, L.Y., Metalsky, G.I., Seligman, M.E.P. (1982). The Attributional Style Questionnaire. *Cognitive Therapy and Research, 6*, 287-299.

Rosenberg, M. (1979). Concerning the Self. New York: Basic Books.

Schumacker, R.E. & Lomax, R.G. (1996). A Beginner's Guide to Structural Equation Modelling. Lawrence Erlbaum: New Jersey.

Simon, G., Von Korff, M. & Barlow, W. (1995). Health care costs of primary care patients with recognized depression. *Archives of General Psychiatry, 52*, 850-856.

Steer, R.A. & Clark, D.A. (1997). Psychometric characteristics of the Beck Depression Inventory-II with college students. *Measurement and Evaluation in Counselling and Development, 30*, 128-136.

Vassend, O., Lian, L. & Andersen, H.T. (1992). Norske versjoner av NEO-Personality Inventory, Symptom Checklist 90-Revised og Giessen Subjective Complaints List. Del 1. *Tidsskrift for Norsk Psykologforening, 29*, 1150-1160.

Vredenburg, K., Krames, L. & Flett, G.L. (1985). Reexamining the Beck Depression Inventory: The long and short of it. Psychological Reports, 57, 767-778.

Watson, D. & Clark, L.A. (1984). Negative Affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin, 96*, 465-490.

Whisman, M.A., Perez, J.E. & Ramen, W. (2000). Factor Structure of the Beck Depression Inventory-II in a student sample. *Journal of clinical psychology, 56*, 545-551.

Wittchen, H.U., Knauper,B.,& Kessler, R.C. (1994). Lifetime risk of Depression. British *Journal of Psychiatry, 165*, 16-22.

Zung, W.W.K. (1965). A Self-rating Depression Scale. *Archives of General Psychology, 12*, 63-70.

Appendix

Tables and figures: **side:**

Measures:

BDI-II; Beck Depression Inventory-II

BAI; Beck Anxiety Inventory

Zung-SDS; Zung Self-Report Depression Scale

SCL-90-R; Symptoms CheckList, Depression/psychotisism subscale randomized items

ASQ; Attributional Style Questionnaire

Table 3: Psychometric characteristics and factor structure on the BDI-II from studies with student samples

	Psychometric characteristics and factor structure				
Study reference	**Sample**	**Mean (SD)**	**Sex diff.**	**Alpha**	**Method/factors**
Beck et al., 1996	120 students	12.6 (9.9)	W>M	.93	Explorative/ 2 factors cognitive-affective/somatic
Steer & Clarke, 1997	160 students	11.9 (8.1)	n.s.*	.89	Explorative/ 2 factors cognitive-affective/somatic
Osman et al., 1997	230 students	10.9 (7.9)	W>M	.90	Confirmational/ 3 factors neg.attitude/perform./somatic
O'hara et al, 1998	152 students	8.8 (8.1)	n.s.*	not rep.	Not rep.
Dozois et al, 1998	1022 students	9.11 (7.6)	n.s.*	.91	Explorative/confirmational 2 factors cognitive-affective/somatic
Whisman et al, 2000	576 students	8.4 (7.2)	not rep.	.89	Confirmational/ 2 factors cognitive-affective/somatic

*Note: no significant difference

Table 5a: means, standard deviation, corrected item-total correlations, skewness og kurtosis
 for BDI-II-items weighted general population sample (N=875)

BDI-II Item/symptom		M	SD	r (item-total)	Skewness	Kurtosis
1.	Sadness	.20	.46	.64	2.4	6.7
2.	Pessimism	.40	.58	.58	1.4	2.1
3.	Past failure	.35	.65	.54	1.7	1.9
4.	Loss of pleasure	.46	.61	.62	1.1	1.2
5.	Guilt feelings	.45	.54	.48	0.6	-0.5
6.	Punishment feelings	.18	.56	.37	3.5	12.8
7.	Self-dislike	.35	.66	.66	2.0	3.9
8.	Self-critical	.32	.61	.62	1.9	3.5
9.	Suicidal thoughts	.08	.29	.46	3.9	18.0
10.	Crying	.26	.68	.47	2.8	7.9
11.	Agitation	.27	.52	.49	1.9	4.1
12.	Loss of interest	.34	.59	.60	1.7	3.1
13.	Indecisiveness	.35	.57	.52	1.6	2.5
14.	Worthlessness	.29	.56	.64	1.9	3.4
15.	Loss of energy	.74	.60	.59	0.3	0.1
16.	Changes in sleep	.72	.82	.46	1.1	0.9
17.	Irritability	.33	.58	.55	1.7	2.6
18.	Appetie changes	.31	.59	.46	2.1	4.7
19.	Concentration	.44	.59	.59	1.0	0.5
20.	Fatigue	.62	.63	.59	0.7	0.3
21.	Loss of sexual interest.	.57	.80	.38	1.3	1.3
	BDI-II scale	8.12	7.51		1.4	2.4

Corrected Cronbach's alpha BDI-II scale= .905

Table 5b: intercorrelations BDI-II items, weighted general population sample (N=875)

BDI-II item/symptom	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. Sadness	1.0																				
2. Pessimism	.47	1.0																			
3. Failure	.45	.45	1.0																		
4. Loss of pleasure	.49	.46	.34	1.0																	
5. Guilt feelings	.34	.31	.38	.28	1.0																
6. Punishment	.33	.24	.35	.20	.22	1.0															
7. Self-dislike	.49	.44	.51	.41	.32	.31	1.0														
8. Self-critical	.48	.40	.46	.42	.43	.25	.55	1.0													
9. Sucidal thoughts	.44	.32	.33	.34	.29	.29	.41	.39	1.0												
10. Crying	.41	.28	.32	.35	.27	.20	.40	.32	.31	1.0											
11. Agitation	.38	.30	.25	.34	.26	.22	.33	.36	.21	.31	1.0										
12. Loss of interest	.43	.38	.32	.53	.30	.17	.40	.42	.35	.28	.38	1.0									
13. Indecisiveness	.39	.37	.29	.39	.29	.20	.37	.33	.21	.27	.30	.37	1.0								
14. Worthlessness	.46	.49	.49	.48	.32	.26	.60	.45	.40	.35	.27	.46	.36	1.0							
15. Loss of energy	.31	.37	.26	.42	.29	.21	.35	.30	.20	.28	.30	.37	.41	.41	1.0						
16. Sleep changes	.31	.26	.20	.27	.23	.19	.27	.27	.18	.21	.29	.32	.25	.27	.35	1.0					
17. Irritability	.38	.30	.26	.34	.32	.16	.41	.42	.20	.28	.42	.43	.30	.35	.39	.32	1.0				
18. Appetite changes	.31	.20	.27	.28	.20	.21	.35	.28	.28	.27	.22	.30	.16	.35	.36	.32	.33	1.0			
19. Concentration	.35	.38	.28	.38	.30	.20	.36	.34	.20	.27	.35	.40	.48	.35	.46	.40	.42	.37	1.0		
20. Fatigue	.34	.30	.27	.43	.30	.22	.34	.33	.24	.27	.32	.38	.38	.33	.61	.40	.29	.33	.49	1.0	
21. Loss of sex int.	.16	.27	.29	.29	.18	.16	.26	.24	.11	.18	.14	.21	.21	.27	.38	.24	.25	.22	.24	.33	1.0

Mean inter-item correlation = .326

Table 8a: means, standard deviations, item-total correlations, skewness and kurtosis for BDI-II-items, student sample (N=303)

BDI-II item/symptom	M	SD	r (item-total)	Skewness	Kurtosis
1. Sadness	.17	.46	.47	2.9	11.0
2. Pessimism	.41	.62	.51	1.5	2.2
3. Past failure	.31	.56	.57	1.6	1.5
4. Loss of pleasure	.28	.50	.43	1.3	0.8
5. Guilt feelings	.51	.57	.44	0.7	0.1
6. Punishment feelings	.18	.52	.33	3.5	14.0
7. Self-dislike	.30	.68	.45	2.4	5.5
8. Self-critical	.32	.61	.33	1.9	3.1
9. Suicidal thoughts	.12	.31	.45	2.3	3.7
10. Crying	.26	.66	.61	2.9	8.5
11. Agitation	.38	.53	.43	1.1	1.0
12. Loss of interest	.24	.48	.46	2.0	4.7
13. Indecisiveness	.45	.69	.37	1.7	3.3
14. Worthlessness	.17	.48	.39	2.8	6.9
15. Loss of energy	.50	.61	.42	0.9	0.3
16. Sleep changes	.69	.69	.52	0.9	1.2
17. Irritability	.27	.50	.31	1.8	3.5
18. Appetite changes	.42	.62	.37	1.5	2.7
19. Concentration	.47	.61	.46	0.9	-0.3
20. Fatigue	.41	.56	.49	0.9	-0.1
21. Loss of sexual interest	.15	.38	.32	2.5	2.7
BDI-II scale	7.12	6.0		1.4	2.7

Corrected Cronbach's alpha BDI-II = .861

Table 8b: intercorrelations BDI-II items, student sample (N=303)

BDI-II item/symptom	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. Sadness	1.0																				
2. Pessimism	.46	1.0																			
3. Failure	.27	.45	1.0																		
4. Loss of pleasure	.30	.24	.29	1.0																	
5. Guilt feelings	.17	.33	.38	.19	1.0																
6. Punishment	.23	.29	.21	.17	.25	1.0															
7. Self-dislike	.25	.32	.41	.21	.20	.25	1.0														
8. Self-critical	.27	.36	.41	.26	.32	.28	.50	1.0													
9. Sucidal thoughts	.35	.33	.38	.23	.24	.34	.25	.36	1.0												
10. Crying	.32	.18	.29	.21	.24	.17	.34	.37	.32	1.0											
11. Agitation	.15	.17	.24	.22	.24	.18	.08	.25	.09	.11	1.0										
12. Loss of interest	.24	.20	.19	.32	.24	.08	.14	.22	.20	.14	.11	1.0									
13. Indecisiveness	.13	.22	.27	.23	.26	.15	.15	.41	.17	.23	.22	.23	1.0								
14. Worthlessness	.34	.43	.43	.22	.26	.13	.40	.43	.33	.28	.20	.20	.23	1.0							
15. Loss of energy	.24	.26	.36	.30	.23	.06	.21	.29	.18	.30	.12	.34	.26	.28	1.0						
16. Sleep changes	.14	.16	.10	.11	.07	.11	.14	.26	.14	.13	.25	.09	.13	.20	22	1.0					
17. Irritability	.32	.20	.17	.18	.15	.11	.18	.25	.12	.20	.16	.33	.20	.26	.25	.10	1.0				
18. Appetite changes	.13	.14	.19	.18	.20	.12	.15	.31	.14	.23	.22	.11	.17	.15	.19	.27	.14	1.0			
19. Concentration	.17	.21	.32	.20	.29	.15	.18	.18	.15	.32	.33	.19	.23	.17	.47	.24	.07	.26	1.0		
20. Fatigue	.22	.20	.27	.23	.20	.05	.17	.27	.13	.26	.25	.28	.24	.20	.55	.28	.24	.29	.40	1.0	
21. Loss of sex int.	.21	.13	.27	.25	.15	.21	.14	.15	.06	.13	.16	.32	.15	.16	.27	.03	.30	.07	.13	.27	

Mean inter-item correlation = .227

Table 6: BDI-II total scores converted to Z-scores, weighted general population sample (N=875)

Sum BDI-II	Z-score	Sum BDI-II	Z-score
0	-1,08	24	2,12
|1	-0,95	25	2,25
2	-0,82	26	2,38
3	-0,68	27	2,52
4	-0,55	28	2,65
5	-0,42	29	2,78
6	-0,28	30	2,92
7	-0,15	31	3,05
8	-0,02	32	3,18
9	0,12	33	3,32
10	0,25	34	3,45
11	0,38	35	3,58
12	0,52	36	3,71
13	0,65	37	3,85
14	0,78	38	3,98
15	0,92	43	4,65
16	1,05		
17	1,18		
18	1,32		
19	1,45		
20	1,58		
21	1,72		
22	1,85		
23	1,98		

Table 4: number of respondents by age and sex, unweighted general population sample (N=875)

	Age				
	<30 years	30-40	45-59	>60	Total
Women	67	223	162	60	512
Men	51	142	108	62	363
Total	118	365	270	122	875

Table 7: Means and standard deviances for total BDI-II score by sex and age group, weighted general population sample (N=875)

	Age				
	<30 years	30-40	45-59	>60 years	Total
Women	10.3 (9.2)	8.7 (8.0)	7.7 (7.9)	8.9 (5.8)	8.9 (7.7)**
Men	8.4 (7.9)	6.2 (6.9)	7.5 (7.6)	7.5 (6.1)	7.3 (7.1)**
Total	9.3 (8.6)	7.4 (7.5)	7.6 (7.7)	8.3 (5.9)	

**Note: significant sex difference ($p < .002$)

Table 9: Groups by mean total BDI-II-score based on Beck et al., 1996 BDI-II manual, weighted general population sample (N=875)

Group (Total BDI-II-score)	Total	%
0-13 (minimal)	711	81,0
14-19 (mild)	93	10,6
20-28 (moderat)	52	6,0
29-63 (severe)	21	2,4

Table 10: means and standard deviation total BDI-II score by sex in the student sample
(N=303)

	BDI-II	Total N
Women	7.4. (5.8)	212
Men	6.3 (6.1)* n.s.	91
Total	7.1 (6.0)	303

*Note: not significant (p>.155, 2-tailed)

Table 11: Groups by mean total BDI-II-score based on Beck et al. (1996) BDI-II manual,

student sample (N=303)

Group (BDI-II-score)	Total	%
0-13 (minimal)	272	89
14-19 (mild)	16	5,8
20-28 (moderat)	12	4,2
29-63 (severe)	3	1,0

Table 12: Bivariate correlations (Pearsons r) between BDI-II and convergent and
discriminant scales, student sample (N=303)

Scale	BDI-II	Zung-SDS	SCL-90 Dep.	BAI	SCL-90 Psyc.
BDI-II	-				
Zung-SDS	.71**	-			
SCL-90 Dep.	.84**	.69**	-		
BAI	.67**	.60**	.69**	-	
SCL-90 Psyc.	.61**	.56**	.72**	.60**	-

** Note: correlation significant .01 level (2-tailed).

Table 13: Means, standard deviation and alpha values for scales in the student sample (N=303)

Scale	Mean	SD	Alpha
SCL-90-depression	6.6	7.4	.89
SCL-90-psychotisism	2.0	3.2	.71
Zung-SDS	34.4	8.0	.85
BAI	5.4	5.9	.87
ASQ (N=249)			
Positive events	4.9	0.6	.82
Internality	4.7	0.7	.49
Stability	5.0	0.8	.61
Globality	4.7	0.9	.67
Importance	5.5	0.8	.67
Negative events	4.2	0.7	.79
Internality	4.2	0.7	.47
Stability	4.2	0.9	.64
Globality	3.9	1.0	.75
Importance	4.7	1.0	.70

Table 14: Mean of mean scores on ASQ subscales by BDI-II group (N=249)

	BDI-II <=13 (N=219)	BDI-II >13 (N=30)	
ASQ			$F_{(1, 247)}$
Positive events			
Internality	4.7 (0.7)	4.3 (0.7)**	9.3, p<.002
Stability	5.1 (0.7)	4.3 (0.7)**	27.4, p<.001
Globality	4.8 (0.9)	4.4 (0.9)	
Importance	5.5 (0.8)	5.5 (0.8)	
Negative events			
Internality	4.2 (0.7)	4.5 (0.8)**	6.7, p<.01
Stability	4.1 (0.7)	4.3 (0.8)	
Globality	3.8 (1.1)	4.2 (0.8)*	4.2, p<.04
Importance	4.7 (1.0)	4.8 (0.7)	

*Note: significant group effect on attribution dimension (.05-level, 2-tailed)
**Note: significant group effect on attribution dimension (.01-level, 2-tailed)

Table 15: goodness of fit indices for the BDI-II factor models, unweighted general population sample (N=875)

Model	df	Chi sq.	CFI	AIC	RMSEA
		Fit index*			
2-factor model	188	746 (p<.001)	.91	833.8	.058
3-factor model	186	674 "	.91	768.4	.055
2. order 3-factor model	188	711 "	.91	800.3	.056

*Note: CFI; Comparative Fit Index, AIC; Akaike Information Criterion, RMSEA; Root Mean Square Error of Approximation.

Table 16: goodness of fit indices for the modified three-factor model for subgroups of the unweighted general population sample (N=875)

Model	df	Chi-sq.	CFI	AIC	RMSEA
		Fit index			
Modified 3-factor model	183	525 (p<.001)	.95	620.3	.046
Women (N=512)		351 "	.96	447.5	.042
Men (N=363)		407 "	.91	503.1	.058
Education 1* (N=435)		391 "	.93	487.9	.044
Education 2* (N=440)		398 "	.93	494.7	.052
Marital status 1* (N=661)		453 "	.94	550.9	.047
Marital status 2* (N=214)		278 "	.94	374.3	.050

*Note: education 1: gymnasium or less, education 2: college/university, marital status 1: married/partner, marital status 2: single/divorced/widow/widower.

Table 17: goodness of fit indices factor models, student sample (N=303)

Model	df	Chi-sq.		CFI	AIC	RMSEA
		Fit index*				
2-factor model	188	363 (p<.001)		.86	451.7	.056
3-factor model	186	351	"	.87	444.3	.054
2. order 3-factor model	188	358	"	.87	447.9	.055
Modified 3-factor model	182	310	"	.90	408.6	.048

*Note: CFI; Comparative Fit Index, AIC; Akaike Information Criterion, RMSEA; Root Mean Square Error of Approximation.
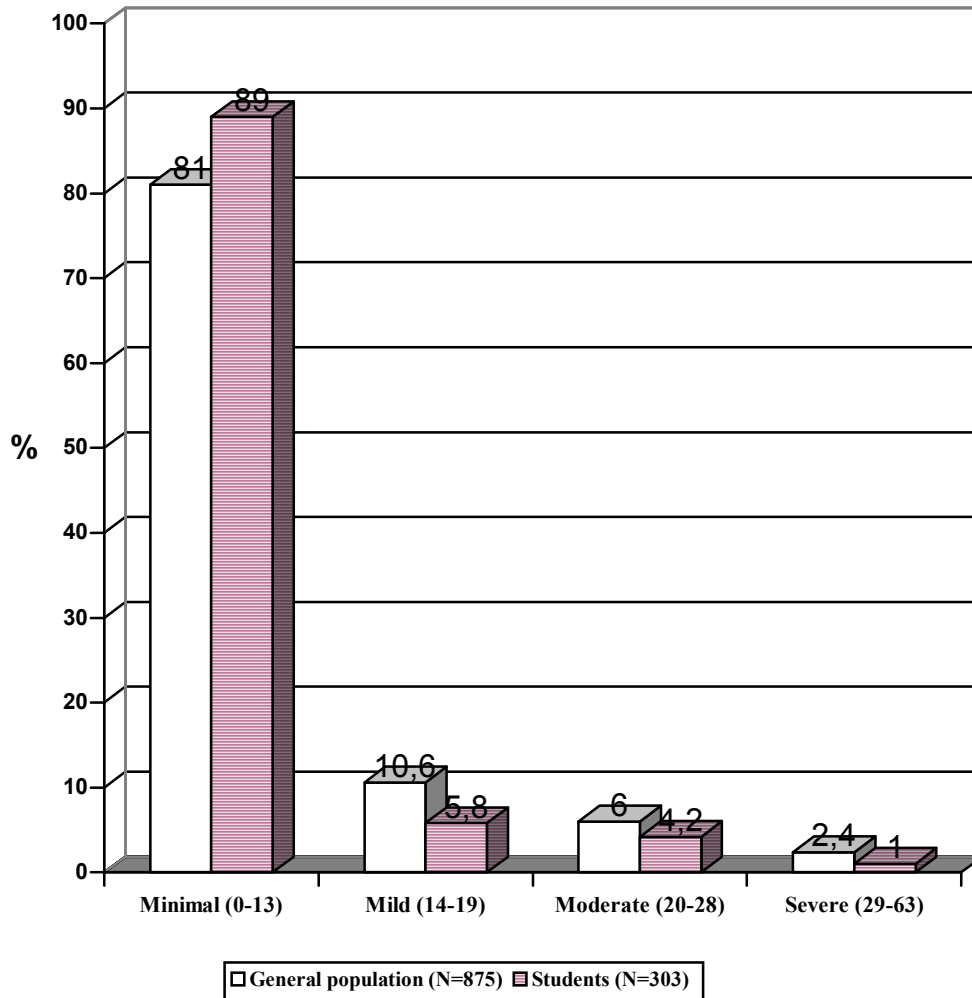
Figure 1: percentage symptom groups **(**Beck et al., 1996) by sample.
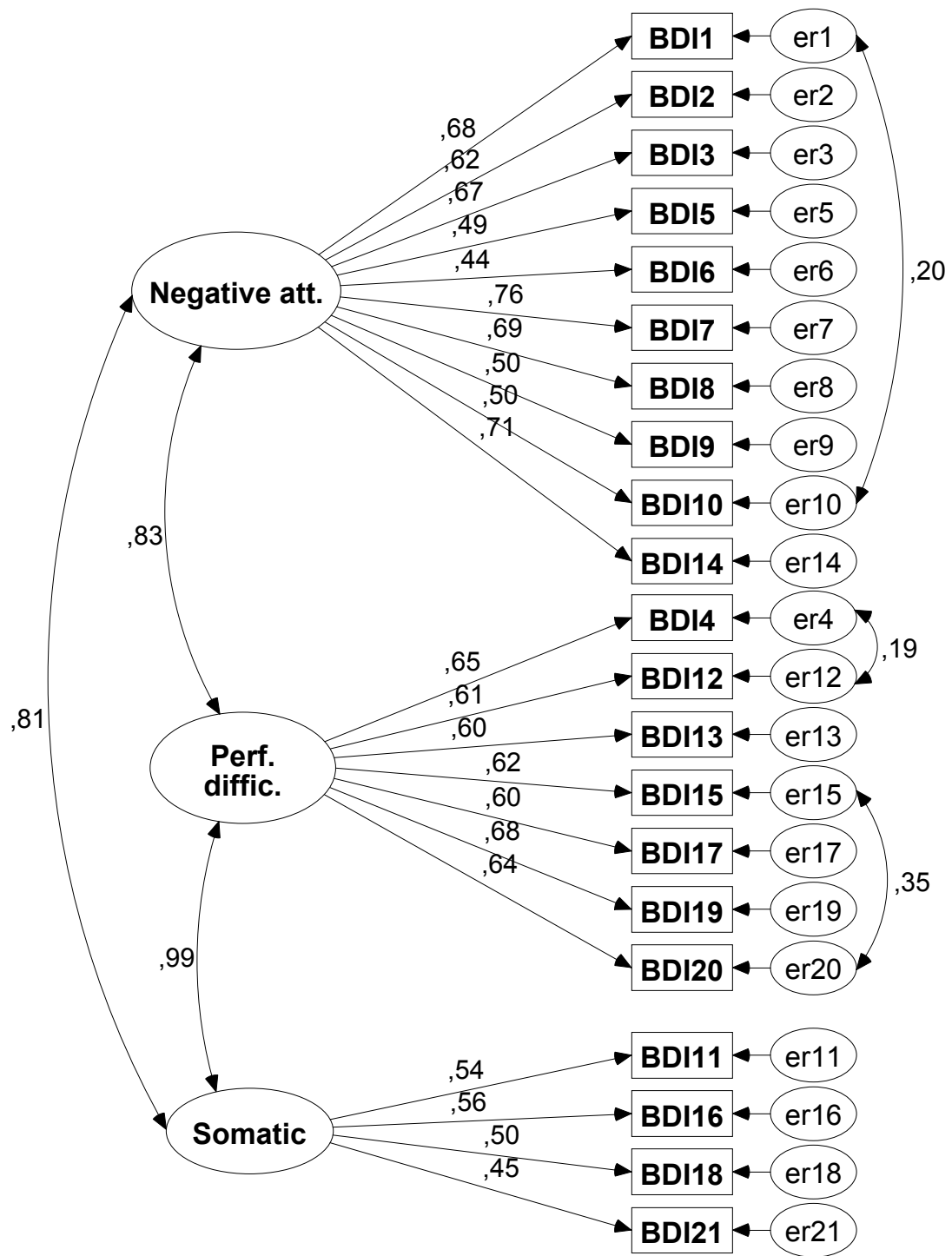
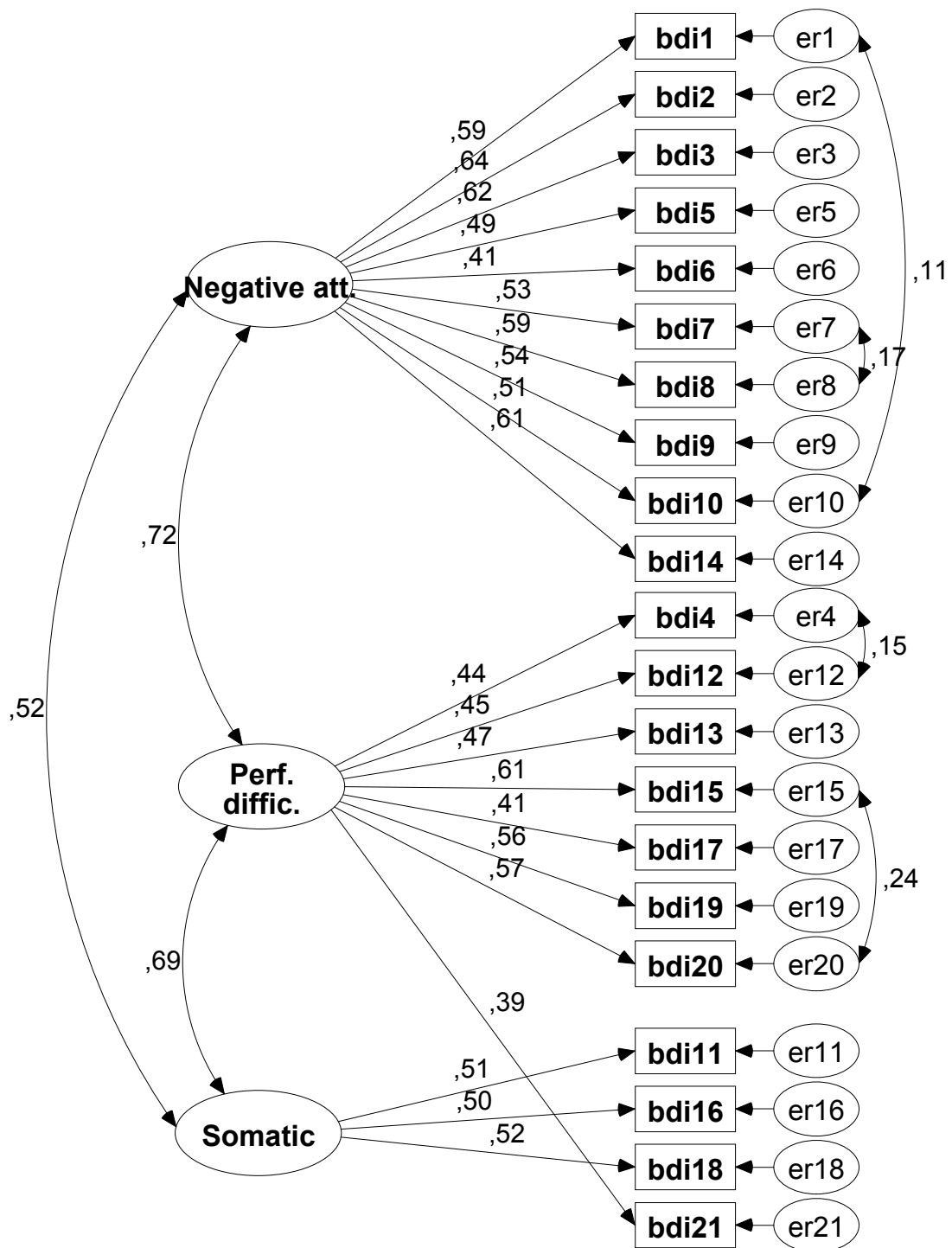**Figure 5**: Modified three-factor model BDI-II for general population data (N=875).

Figure 6: Modified three-factor model BDI-II for student data (N=303).