International Journal of Mental Health Nursing (2007) *16*, 108–115 doi: 10.1111/j.1447-0349.2007.00453.x

FEATURE ARTICLE

Measuring melancholy: A critique of the Beck Depression Inventory and its use in mental health nursing

Brad Hagen
School of Health Sciences, The University of Lethbridge, Lethbridge, Alberta, Canada

ABSTRACT: *The Beck Depression Inventory (BDI) is one of the most commonly used depression measurement instruments. Mental health nurses often utilize the BDI to assess the level of depression in clients, and to monitor the effectiveness of treatments such as antidepressants and electroconvulsive therapy. Despite the widespread use of the BDI in both clinical practice and research, there is surprisingly little nursing literature critically examining the BDI or its use by mental health nurses. This paper reviews the origins, purpose, and format of the BDI, discusses some of the strengths and limitations of the BDI, and concludes with some implications for mental health nursing.*

KEY WORDS: *Beck Depression Inventory*, *depression*, *mental health*, *psychiatric nursing*, *psychometrics*.

Mental health nurses are often encouraged to use psychiatric/psychological measurement instruments in their nursing practice, and the Beck Depression Inventory (BDI; Beck *et al*. 1996) is one of the most commonly used instruments that mental health nurses are likely to encounter and use in their practice (Demyttenaere & De Fruyt 2003). The majority of current psychiatric nursing textbooks discuss the BDI, and in one commonly used textbook, the BDI is described as '. . . a quick but reliable and valid measure of the extent to which depression may be present' (Kneisl *et al*. 2004; p. 169). Indeed, in the first author's own local health region, where he supervises nursing students in mental health clinical settings, nurses commonly use the BDI to assess the level of depression in patients, and to monitor the effectiveness of treatments such as antidepressants and electroconvulsive therapy.

Yet despite the common use of the BDI by mental health nurses, there is little or no nursing literature critically examining the BDI, or its use by mental health nurses. Therefore, the purpose of this paper is to provide a critical discussion of the BDI, and its use by mental health nurses. To this end, the author will briefly review the origins, purpose, and format of the BDI, discuss some of the strengths and limitations of the BDI, and conclude with some implications for mental health nursing.

ORIGINS, PURPOSE, AND FORMAT OF THE BDI

As Demyttenaere and De Fruyt (2003) have described in their review of depression rating scales, depression rating scales were first developed in the late 1950s as part of the overall psychopharmacology revolution, whereby psychological theories of depression gave way to commercially driven biochemical theories of depression. They go on to note that '. . . with the advent of antidepressant drugs, rating scales were needed to measure the severity of the depressive disorder and the changes during therapy' (Demyttenaere & De Fruyt 2003; p. 61), and there are now more than 100 depression rating scales in existence. While there are many available depression rating scales, the BDI, a self-administered scale first developed in 1961

Correspondence: Brad Hagen, School of Health Sciences, The University of Lethbridge, 4401 University Drive, Lethbridge, AB T1K 3M4, Canada. Email: brad.hagen@uleth.ca
Brad Hagen, RN, PhD.
Accepted May 2006.

(Beck *et al*. 1961), has risen to prominence as one of the most widely used depression rating scales (Dozois & Covin 2004).

Beck *et al*. (1961) developed the original BDI as a 21-item inventory to measure the severity of depressive symptoms, and based the 21 items on Beck's observations of the symptoms and attitudes of depressed persons seen in the context of therapy. Beck has claimed that the BDI does not reflect any particular theory of depression, and merely reflects the observed symptoms of persons who are depressed (Beck *et al*. 1996). While the first version of the BDI was shown to be reasonably robust in terms of psychometric properties, increasing concerns were raised concerning the instrument's validity vis-à-vis the DSM-IV standard for diagnosing depression (American Psychiatric Association 1994). Consequently, Beck *et al*. created a second revised version of the BDI (BDI-II) in 1996 (Beck *et al*. 1996). The main changes made to develop the BDI-II primarily reflected increased compatibility with the DSM-IV, and included the changing of certain items, dropping of other items, and changes to certain response options and time frames (Beck *et al*. 1996; Dozois *et al*. 1998). From this point forwards in this manuscript, the authors will use the term 'BDI' to refer to this most recent version of the instrument, the BDI-II.

The BDI is typically self-administered, requires only about 5–10 min to complete, and can be used with persons aged 13 years and up (Dozois & Covin 2004). Each one of the 21 items in the BDI is rated on a scale of 0–3, and scores from all items are tallied to obtain a total possible score, ranging from 0 and 63, with higher scores reflecting greater severity of depressive symptomatology. Scores between 0 and 13 are interpreted as 'minimal' depression, scores between 14 and 19 as 'mild' depression, scores of 20–28 as 'moderate' depression, and scores of 29–63 as 'severe' depression (Beck *et al*. 1996; Dozois & Covin 2004). Interestingly, it appears that with the possible exception of a score of 0, there are no score grouping to be interpreted as 'no depression'.

STRENGTHS OF THE BDI

Strengths of the BDI include the ease of administration and scoring of the BDI, its widespread use, and the results of psychometric testing of the reliability and validity of the BDI.

Ease of administering and scoring the BDI

One of the principle advantages of the BDI is its ease of administration and scoring (Dozois & Covin 2004). Indeed, the BDI generally only takes less than 10 min to complete, and is easily scored and interpreted. Consequently, the BDI has become one of the most widely used psychological tests, has been translated into many languages, and has been employed in more than 2000 empirical studies (Barroso & Sandelowski 2001; Dozois & Covin 2004; Richter *et al*. 1998).

Psychometric testing of the BDI

Reliability of the BDI

Although more reliability testing has been completed on the original BDI than the BDI-II, both are considered to be generally quite reliable (Dozois *et al*. 1998; Richter *et al*. 1998). The original manual for the BDI-II reported high internal consistency, with a coefficient alpha of 0.93 for college students, and 0.92 for psychiatric outpatients (Beck *et al*. 1996). More recently, Dozois and Covin (2004) reviewed 13 studies reporting reliability data on the BDI-II since 1996, and reported an average coefficient alpha of 0.91. Less information is available on the test–retest reliability of the BDI-II, although the original manual reports a 1-week test–retest reliability coefficient of 0.93 with 26 psychiatric outpatients (Beck *et al*. 1996). As Dozois and Covin (2004) have cautioned, however, test–retest reliability is difficult to interpret on a measure that is supposed to both reliably measure depression *and* detect changes in depression due to treatment. For example, at least one group of researchers have suggested that the BDI may not be reliable for longer periods of time in non-clinical samples, after finding that BDI scores declined by 40% over 2 months in a non-clinical sample (Ahava *et al*. 1998). Such a significant downward drift in BDI scores in non-clinical samples clearly poses a threat to the instrument's ability to reliably detect changes in depression due to treatment alone.

Validity of the BDI

Dozois and Covin (2004) have asserted that while the BDI is comparable to the original BDI in terms of reliability, the BDI-II is '. . . a clearly superior instrument in terms of its validity' (p. 53). Such claims for the higher validity of the BDI-II are made on a number of levels. To begin with, the content validity and the face validity of the BDI-II are argued to be very high, because the items in the BDI-II now closely mirror the standard DSM-IV diagnostic criteria for depression (Dozois & Covin 2004; Richter *et al*. 1998). The convergent validity of the BDI-II has also been reported, and the BDI-II appears to correlate fairly well with other depression rating scales,

such as the original BDI-I ($r = 0.93$), the Hamilton Rating Scale for depression ($r = 0.71$), and the Beck Hopelessness Scale ($r = 0.68$) (Dozois & Covin 2004).

The level of discriminate validity for the BDI is less clear. For example, while Richter *et al*. (1998) have concluded that the BDI '. . . discriminates reliably between depressives and non-depressives' (p. 162), Dozois and Covin (2004) came to the opposite conclusion, stating that the BDI-II does *not* differentiate well between depressed and non-depressed persons. Furthermore, it has been noted that the BDI-II correlates highly with other measures of anxiety, and may not be able to reliably distinguish between depression and other affective states such as anxiety (Dozois & Covin 2004).

Finally, it has been suggested that the BDI-II contains a reasonably stable factor structure (Beck *et al*. 1996; Dozois & Covin 2004). When the BDI-II was first released, Beck *et al*. (1996) reported a two-factor solution: somatic-affective and cognitive symptoms within a psychiatric outpatient sample, and cognitive-affective and somatic symptoms with a college student sample. Other researchers have since found generally similar two-factor structures in other studies using college students (Dozois *et al*. 1998) and primary care medical patients (Arnau *et al*. 2001). It should be noted, however, that in their review of the BDI, Richter *et al*. (1998) concluded that the tacit factorial validity of the BDI is in fact controversial, and that subtle but possibly important differences exist in the factor structure of the BDI, depending upon the kinds of subjects that complete the BDI.

In summary, the main support for the BDI appears to lie in its ease of use, widespread utilization, very good internal reliability, high content validity when compared with the DSM-IV criteria for depression, good convergent validity with other similar depression rating scales, and a somewhat stable factor structure.

LIMITATIONS OF THE BDI

While the BDI is well-known and widely used by mental health nurses, and while the BDI has several strengths, there is little critical discussion in the nursing literature of some of the potential limitations of the use of the BDI in general, or by mental health nurses in particular. Some of the potential limitations of the BDI include: issues related to norms (including potential bias issues); problems with the wording, ordering, and weighting of the BDI items; potential gender biases; theoretical issues with the BDI; potentially inappropriate uses of the BDI; and validity issues related to the DSM-IV criteria for depression, upon which the BDI is based.

Norms and bias issues

The BDI has no actual large population norms per se, so it is difficult to determine if any given individual's level of depression, as determined by the BDI, is 'normal' in any sense of the word. Instead, the interpretation of the BDI is referenced to criterion based on the original standardized sample of 500 persons (317 women and 183 men) in the Eastern United States (Beck *et al*. 1996). Based on this sample, the authors of the manual for the BDI-II offered cut-off score criterion or guidelines to distinguish between minimal, mild, moderate, and severe amounts of depression. However, while the total possible scores range from 0 to 63, the scoring of the scale is very 'bottom heavy'. That is, the mean score for severely depressed persons in the standardized sample (32.96) is approximately half-way along the range of total possible scores, and anyone who scores anywhere from 29 to 63 is considered to be 'severely' depressed (Beck *et al*. 1996).

As several authors have noted, the original sample upon which the BDI-II was standardized was predominantly Caucasian, and is greatly misrepresentative of the US population at large (Dozois & Covin 2004; Richter *et al*. 1998). Obviously, this kind of sample also renders the BDI generally misrepresentative of other countries and cultures (Dozois & Covin 2004), and fails to capture the many different cultural factors influencing how depression is experienced by different ethnic and cultural groups (Falicov 2003). Finally, women tend to score higher on the BDI than men (Beck *et al*. 1996) and items on the BDI such as 'crying' may contain a gender bias, and may hold very different meanings for men as opposed to women (Barroso & Sandelowski 2001).

Item-related issues

There are also several problems with the way that items contained in the BDI are worded, ordered, and weighted. To begin with, several authors (Barroso & Sandelowski 2001; Demyttenaere & De Fruyt 2003; Richter *et al*. 1998) have noted that the BDI item response options, most of which contain some combination of negatively and positively worded options, can be very confusing and misleading for persons taking the BDI. In addition, the responses are only ordinal-level data, with unequal intervals between options, yet are tallied up, analysed, and reported as if they are ratio-level data (Burns & Grove 2001). There is also a tendency for responses on each item to score quite low. That is, although potential scores for each item range from 0 to 3, studies in non-clinical (student) samples typically report average scores below 1, and even psychiatric samples mean item scores rarely exceed values of 2 (Richter *et al*. 1998).

The BDI also has an obvious consistency in the ordering of responses, as each response is ordered from least to most depressed. Therefore, several authors have noted that the obvious ordering of answers in the BDI may lead to responses reflective of faking, social desirability, and/or defensiveness, as opposed to depression per se (Barroso & Sandelowski 2001; Dozois & Covin 2004). This problem is compounded by the BDI's high face validity, which makes it easy for subjects to guess which of the items are reflective of greater or lesser depression.

Finally, there is the problem of item weighting within the BDI. As Healy (1997) has noted, the creators of the BDI arbitrarily decided that the total BDI depression score would be generated by simply adding the individual scores for each of the 21 items. This kind of scoring system raises the issue of whether or not it is legitimate to simply add dissimilar items to produce a total score, and subsequently assume that the total score actually 'means' something (this process is called reification, and shall be addressed further in this paper). This scoring process also raises the issue of whether or not each added item in the BDI should have equal weighting in relation to the other items. Healy (1997) has commented on this problem, and has noted that '. . . many clinicians had difficulties with the idea that items of very different meaning could simply be summed. Should early morning awakening be counted in the same balance as guilt or suicidality?' (p. 98). Yet despite such criticisms, all items of the BDI continue to be treated as if they are of equal importance in determining a person's level of depression, and the creators of the BDI have offered no justification for such a stance.

Theoretical issues with the BDI

There are a number of important theoretical limitations with the BDI as well. First and foremost is the problem with the supposedly atheoretical nature of the BDI. That is, although the creators of the BDI maintain that the BDI merely reflects the symptoms and attitudes typically found in persons with depression – and does not reflect any theoretical assumptions about depression (Beck *et al.* 1996; Dozois & Covin 2004) – other authors have challenged this claim of theoretical neutrality. Demyttenaere and De Fruyt (2003), for example, have noted that the BDI clearly reflects a distinctly cognitive–behavioural perspective. This perspective is not surprising, given that Beck *et al.* were primarily responsible for the creation of cognitive therapy. Healy (1997) has also observed that it is probably more than coincidence that the BDI is particularly well-suited for evaluating cognitive–behavioural therapy, and that it would be very difficult for a person who has gone through cognitive therapy not to recognize many of the terms and language used in the BDI. Therefore, the assertion that the BDI is theoretically neutral of bias-free is simply not true, nor should this necessarily be surprising. As Jensen and Hoagwood (1997) have emphasized:

> . . . it should be noted that *all* clinicians – indeed, all human beings – bring theory-laden perspectives and conceptual filters to their assessment and diagnostic approaches with a given patient. They differ principally in the explicitness, rigidity and awareness of their biases (p. 235).

Perhaps one of the most important source of bias found within the BDI is reflected in what the architects of the BDI chose *not* to include as items in the tool. For example, the BDI focuses exclusively on negative symptomatology – such as sadness, guilt, and feeling like a failure – and no positive experiences symptoms are included, despite research suggesting that positive mood may well be superior to negative mood in predicting outcomes from depression (Demyttenaere & De Fruyt 2003). In addition, the creators of the BDI chose to disregard large areas of interpersonal functioning, and many of the factors which determine quality of life for individuals (Healy 1997). Finally, Beck *et al.* have selected items for the BDI that clearly reflect a theoretical stance whereby the problem (i.e. depression) is seen to lie *within* the individual. By focusing exclusively on symptoms or problems inside the person, the BDI explicitly disregards all the multitude of factors and problems external to the individual that may be clearly impacting his or her level of depression, such as unemployment, discrimination and/or domestic violence (Crowe 2000; Jensen & Hoagwood 1997).

Lastly, the BDI exhibits the theoretical problem of *reification*, or the tendency to view abstract concepts as actual entities. That is, the creators of the BDI would have us believe that the simple process of adding up the answers to 21 questions about various symptoms and attitudes allows us to measure, with a single number, the quantity of a reliably identifiable 'thing' called depression, as if we were measuring the weight or height of an individual. Yet as Gould (1996) has pointed out, measuring and reifying such concepts as 'depression' and 'intelligence', as if they have a definite existence of their own, can be very misleading. Not only can such reification oversimplify complex and multifaceted experiences like depression, but such reification also disregards the large extent to which such concepts are socially created and defined, and fail to actually reflect any clearly tangible and

unambiguous entities, such as height, weight, or hypertension (Sarbin 1997).

Inappropriate uses of the BDI

Additional concern has also been raised regarding the use of potentially controversial 'spin-offs' of the BDI, as well as the manner in which the BDI is being increasingly used to 'screen' for depression. With regards to spin-offs, Dozois and Covin (2004) have noted the increasing use of controversial spin-offs of the BDI, such as the 'short form' of the BDI (Furlanetto, Mendlowicz & Romildo 2005). The increasing use of such spin-offs, because of their comparative lack of psychometric testing and established norms, only further compounds the overall issues of reliability and validity of the BDI and other depression rating scales.

Furthermore, increasing numbers of clinicians, including nurses, are beginning to use the BDI as a depression 'screening tool', particularly now that the BDI-II closely mirrors the DSM diagnostic criteria for depression (Lasa *et al*. 2000). However, despite the fact that the creators of the BDI specifically specified that the BDI was *not* to be used as a diagnostic tool, and was to only be used as a measurement of depressive symptom severity (Beck *et al*. 1996; Dozois & Covin 2004), the demarcation lines between *measuring* symptoms of depression, *diagnosing* depression, and *screening* for depression have never been clear, and are becoming even less clear. In particular, it was never conceptually clear why the BDI – an instrument apparently able to measure the *quantity* of depression – could not actually determine the *presence* of depression or not (i.e. diagnosis it), particularly when so many of the BDI items used to measure depression symptoms were so similar to the same DSM-IV diagnostic criteria for depression. Despite this logical inconsistency, the BDI is increasingly being used not only to measure depression, but to detect or diagnosis it as well (Lasa *et al*. 2000), despite a lack of clear validation for doing so (Beck *et al*. 1996).

Validity issues and the DSM

When Beck *et al.* attempted to increase the validity of the BDI by making the BDI-II more closely mirror DSM-IV diagnostic criteria for depression (Beck *et al*. 1996), they also further reinforced the common assumption that the DSM-IV offers the most valid definition and description of the experience of depression. By doing so, however, they not only overlooked considerable criticism of the way that the DSM-IV authors categorize depression, but also inherited many of the limitations of the DSM-IV description of depression (Beutler & Malik 2002; Crowe 2000; Donald 2001; Eriksen & Kress 2005; Jensen & Hoagwood 1997; Sarbin 1997). Therefore, any examination of the limitations of depression scales like the BDI must also include an examination of the limitations of the DSM-IV criteria. These limitations include issues of reliability and validity of the DSM-IV, and issues of DSM-IV value judgements and biases.

Issues of the reliability and validity of the DSM-IV

While many mental health clinicians simply take the reliability and validity of the DSM-IV system for granted, closer examination often finds the reliability and validity of the DSM-IV wanting. In fact, the reliability of the diagnosis of major depression is quite poor, and researchers have reported kappa coefficients for the diagnosis of depression as low as 0.25 (Parker 2005). As Beutler and Malik (2002) have observed, this inadequate level of diagnostic reliability is not surprising, given the ambiguous and complex set of guidelines that the authors of the DSM-IV created to diagnosis depression. In fact, the DSM-IV criteria for depression literally allow for several hundred possible different patterns or clusters of symptoms, all of which can still all meet the DSM-IV diagnostic criteria for depression (American Psychiatric Association 1994).

Given the myriad of symptom patterns which can qualify for a DSM-IV diagnosis of depression, the DSM-IV diagnosis of depression suffers not only from reliability problems, but from considerable validity problems as well. For example, all forms of depression share great overlap with numerous other psychiatric diagnosis contained within the DSM-IV, and depression is found to be comorbid in 60% of general psychiatric patients, and in 40% of patients diagnosed with anxiety disorders (Beutler & Malik 2002). In addition, another key indicator of diagnostic validity – that different diagnoses would respond differently and predictably to prescribed treatments – is also lacking. That is, both the natural and treatment histories of persons with a diagnosis of depression are notoriously hard to predict (Parker 2005). Furthermore, a wide variety of treatments with entirely different theoretical explanations – including antidepressants, different forms of counselling, electroconvulsive therapy, St. John's Wort, exercise, and placebo – all have nearly identical efficacy levels, seriously challenging the diagnostic construct of depression implicitly contained within the BDI (Parker 2005).

Value judgements and biases within the DSM

As opposed to other medical diagnoses, the diagnoses contained within the DSM-IV (including the diagnosis of

depression) are not arrived at with the aid of laboratory tests or diagnostic imaging, but rely instead on a clinician's judgement as to whether or not certain behaviours and symptoms, such as feelings of guilt or loss of concentration, are present in certain prescribed patterns for certain prescribed periods of time (Eriksen & Kress 2005). These prescriptions for what constitutes a diagnosis (or disorder) are in turn arrived at by consensus by committees and panels of psychiatric experts associated with the American Psychiatric Association (1994).

Numerous authors have challenged this DSM diagnosis-by-consensus process, claiming that the process reflects not so much a scientific and objective process, but a process whereby the values and biases of the privileged few comprising 'expert consensus' panels become embedded in our society's definitions of mental disorders (Beutler & Malik 2002; Eriksen & Kress 2005; Jensen & Hoagwood 1997; Kutchins & Kirk 1997; Sarbin 1997). Female scholars in particular (Caplan 1995; Crowe 2000; Russell 1986) have noted the preponderance of upper-middle and upper class men in DSM diagnostic expert committees, and have suggested that Western, male, and upper/middle class values strongly influence decisions regarding diagnoses such as depression, and how such diagnoses are applied. For example, using the standard DSM-IV criteria for depression, twice as many women are diagnosed with depression as men (Kuehner 2003), yet the role that many contextual factors – such as gender discrimination in society or the higher rates of sexual abuse and assault in girls and women – are rarely taken into account when diagnosing or measuring depression (Whitfield 2003).

This disregard of contextual factors reflects another bias inherent within the DSM-IV diagnosis of depression, the notion that mental disorders are located *within* individuals. This tendency to locate mental disorders and problems inside individuals has important implications, as it can easily direct clinicians' attention away from the social context of mental health issues. That is, numerous authors have argued forcibly that it is equally plausible – and perhaps more appropriate – to suggest that it may well be our families, communities, and societies that deserve such labels as 'depressed', 'disordered' or 'mentally ill', as opposed to individual persons (Crowe 2000; Jensen & Hoagwood 1997; Russell 1986; Sarbin 1997; Whitfield 2003). For example, a woman suffering from domestic violence and seeking assistance from the mental health system is likely to receive a psychiatric diagnosis of depression and/or post-traumatic stress disorder, and may be given instruments like the BDI to determine the extent of her 'disorder'. Yet the real source of the woman's problems – the perpetrator of the violence towards her – is typically given no corresponding psychiatric diagnosis (Eriksen & Kress 2005).

SUMMARY AND CONCLUSIONS

In summary, it has been shown that the BDI was created in the historical context of the rise of psychopharmacology and DSM nosological classification systems in the mental health care system, and that the BDI has come to be very widely used in research and practice for measuring the symptoms of depression, and more recently, for the screening of depression. Consequently, mental health nurses are often in a position of administering and interpreting the BDI for a variety of clients and reasons, in a variety of mental health settings.

The BDI demonstrates high internal consistency, and advocates of the BDI claim that it has high validity, particularly with regard to high construct validity compared with the DSM-IV diagnostic criteria for major depression. However, despite widespread use of the BDI and numerous claims for the strong psychometric properties of the BDI, the BDI has several significant limitations that mental health nurses need to consider. These limitations include the numerous potential cultural, gender and/or theoretical biases inherent in the BDI, the increasing use of poorly researched spin-offs of the BDI, using the BDI for purposes for which it was never intended, and the many limitations of the BDI associated with its adoption of the DSM-based description of depression.

These limitations of the BDI pose numerous important implications for mental health nurses. To begin with, nurses using the BDI to arrive at 'depression scores' for clients need to remember that arriving at such scores is not as simple as measuring someone's blood pressure or weight. Instead, instruments like the BDI reflect a subtle yet important reification of an abstract concept (depression) which has largely been socially constructed, and remains quite idiosyncratic and elusive. Furthermore, the fact that the BDI is so simple and expeditious to use tends to effectively conceal the many assumptions and values that all such mental health instruments contain.

Mental health nurses may also wish to consider that the BDI conceptually locates the source of the problem of depression or 'disorder' within the individual, and may blind us to everything the BDI *is not* measuring that occurs within the larger context of the individual. As Crowe (2000) aptly reminds us, the BDI, based on DSM-IV diagnostic criteria:

> . . . provides little information regarding the gendered, cultural and class aspects of the individual's experience: it provides no information upon which a nurse-patient relationship can be developed, nor what may be useful in facilitating the individual's courage, hope and energy to cope with their mental distress (p. 584).

Therefore, if mental health nurses chose to use tools like the BDI in their care of clients who struggle with depression, they are encouraged to do so with a thorough understanding of the strengths and limitations of the BDI, and a critical awareness of how the tool's underlying values and assumptions may impact the care of that individual, and the nurse–client relationship.

ACKNOWLEDGEMENT

The author would like to acknowledge and thank Dr Robert Williams, who offered a valuable critique of this manuscript.

REFERENCES

Ahava, G. W., Lannone, C., Grebstein, L. & Schirling, J. (1998). Is the Beck Depression Inventory reliable over time? An evaluation of multiple test–retest reliability in a non-clinical college student sample. *Journal of Personality Assessment*, *55*, 224–233.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. Washington, DC: Author.

Arnau, R. C., Meagher, M. W., Norris, M. P. & Bramson, R. (2001). Psychometric evaluation of the Beck Depression Inventory II with primary care medical patients. *Health Psychology*, *20* (2), 112–119.

Barroso, J. & Sandelowski, M. (2001). In the field with the Beck Depression Inventory. *Qualitative Health Research*, *11* (4), 491–504.

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

Beck, A. T., Steer, R. A. & Brown, G. K. (1996). *Beck Depression Inventory Manual*, 2nd edn. San Antonio, TX: Psychological Corporation.

Beutler, L. E. & Malik, M. L. (2002). Diagnosis and treatment guidelines: The example of depression. In: L. E. Beutler & M. L. Malik (Eds). *Rethinking the DSM: A Psychological Perspective* (pp. 251–278). Washington, DC: American Psychological Association.

Burns, N. & Grove, S. K. (2001). *The Practice of Nursing Research*, 4th edn. Philadelphia, PA: W.B. Saunders.

Caplan, P. (1995). *They Say You're Crazy: How the Most Powerful Psychiatrists Decide Who's Normal*. New York: Perseus Books.

Crowe, M. (2000). Psychiatric diagnosis: Some implications for mental health nursing care. *Journal of Advanced Nursing*, *31* (3), 583–589.

Demyttenaere, K. & De Fruyt, J. (2003). Getting what you ask for: On the selectivity of depression rating scales. *Psychotherapy and Psychosomatics*, *72*, 61–70.

Donald, A. (2001). The Wal-Marting of American psychiatry: An ethnography of psychiatric practice in the late 20th century. *Culture, Medicine and Psychiatry*, *25*, 427–439.

Dozois, D. J. A. & Covin, R. (2004). The Beck Depression Inventory-II (BDI-II), Beck Hopelessness Scale (BHS), and Beck Scale for Suicide Ideation (BSS). In: M. Hersen (Series Ed.), D. L. Segal and M. Hilsenroth (Volume Eds). *Comprehensive Handbook of Psychological Assessment: Volume 2 Personality Assessment and Psychopathology* (pp. 50–69). New York, NJ: Wiley.

Dozois, D. J. A., Dobson, K. S. & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory – II. *Psychological Assessment*, *10* (2), 83–89.

Eriksen, K. & Kress, V. E. (2005). *Beyond the DSM Story: Ethical Quandaries, Challenges and Best Practices*. Thousand Oaks, CA: Sage.

Falicov, C. J. (2003). Culture, society and gender in depression. *Journal of Family Therapy*, *25*, 371–387.

Furlanetto, L. M., Mendlowicz, M. V. & Romildo, B. J. (2005). The validity of the Beck Depression Inventory – Short Form as a screening and diagnostic instrument for moderate and severe depression in medical inpatients. *Journal of Affective Disorders*, *86* (1), 87–91.

Gould, S. J. (1996). *The Mismeasure of Man*. New York: W.W. Norton.

Healy, D. (1997). *The Antidepressant Era*. Cambridge: Harvard University Press.

Jensen, P. S. & Hoagwood, K. (1997). The book of names: DSM-IV in context. *Development and Psychopathology*, *9*, 231–249.

Kneisl, C., Wilson, H. & Trigoboff, E. (2004). *Contemporary Psychiatric-Mental Health Nursing*. Upper Saddle River: Pearson Prenctice Hall.

Kuehner, C. (2003). Gender differences in unipolar depression: An update of epidemiological findings and possible explanations. *Acta Psychiatr Scandinavica*, *108*, 163–174.

Kutchins, H. & Kirk, S. A. (1997). *Making Us Crazy: DSM: The Psychiatric Bible and the Creation of Mental Disorders*. New York: Free Press.

Lasa, L., Ayuso-Mateos, J. L., Vazquez-Barquero, J. L., Diez-Manrique, F. J. & Dowrick, C. F. (2000). The use of the Beck Depression Inventory to screen for depression in the general population: A preliminary analysis. *Journal of Affective Disorders*, *57*, 261–265.

Parker, G. (2005). Beyond major depression. *Psychological Medicine*, *35*, 467–474.

Richter, P., Werner, J., Heerlein, A., Kraus, A. & Sauer, H. (1998). On the validity of the Beck Depression Inventory: A review. *Psychopathology*, *31* (3), 160–168.

Russell, D. (1986). Psychiatric diagnosis and the oppression of women. *Women and Therapy*, *5* (4), 83–98.

Sarbin, T. R. (1997). On the futility of psychiatric diagnostic manuals (DSMs) and the return of personal agency. *Applied and Preventive Psychology*, *6*, 233–243.

Whitfield, C. L. (2003). *The Truth About Depression: Choices for Healing*. Deerfield Beach: Health Communications.