CRITERION, CONVERGENT, AND DISCRIMINANT VALIDATION OF THE HUBLEY DEPRESSION SCALE FOR OLDER ADULTS (HDS-OA)

by

Sherrie Myers

B. Eng., Memorial University of Newfoundland, 1996

A THESIS SUBMITED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

The Faculty of Graduate Studies

(Counselling Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

October 2009

Abstract

There is a current need for an accurate screening measure for depression in older adults in both clinical and research settings. The Hubley Depression Scale for Older Adults (HDS-OA; Hubley, 1998) is a short 16-item depression screen designed for the elderly that is based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). It uses a yes/no response format, large font size, and reminders of the reference period. Currently, it lacks sufficient psychometric evidence to support its use. The only other depression screen designed for older adults is the Geriatric Depression Scale (GDS; Yesavage et al., 1983). The 30-item GDS and the shorter form GDS-15 (Sheikh & Yesavage, 1986) also use a yes/no response format but are not based on DSM-IV criteria for depression.

The purpose of the study was to examine and compare the psychometric properties of the HDS-OA, GDS, and GDS-15 using a sample of 18 depressed and 18 non-depressed older adults. Validation evidence was presented to support inferences made from the HDS-OA total score, and the results of the measures were compared to determine if the HDS-OA performed as well or better than the GDS and GDS-15. Score reliability was examined through a measure of internal consistency. Construct validity was examined by correlating depression measure scores to provide evidence of convergent validity and through correlations with anxiety, cognitive status, and self-rated health scores to examine evidence of discriminant validity. Criterion-related validity was examined through differences in depressed and non-depressed group scores and levels of sensitivity and specificity found at an optimum cut score.

The study findings revealed high internal consistency for the HDS-OA, evidence of convergent validity with the GDS and GDS-15, evidence of discriminant validity when correlated with anxiety, cognitive status, and self-rated health scores, and a significant difference between group scores, which attributes to the ability of the HDS-OA to differentiate between individuals with depression and those without. In these analyses, the HDS-OA performed as well as the GDS and GDS-15. The HDS-OA revealed higher sensitivity than the GDS and GDS-15 though, indicating slightly better performance in accurately identifying individuals with depression.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to thank Dr. Anita Hubley, my research supervisor, for her guidance, encouragement, and constant availability throughout this study. As well, I would like to express my gratitude to Dr. Zumbo and Dr. Ishiyama, members of my research committee, for their input, guidance, and support during the development of my research proposal.

A very special thank you is also extended to staff members of the following health care units in the Greater Vancouver area who assisted in subject recruitment: the Geriatric Psychiatry Outpatient Program and Short-Term Treatment and Assessment unit at Vancouver General Hospital, the Geriatric Outpatient Clinic at Mount Saint Joseph Hospital, the Elder Care Program at St. Paul's Hospital, the North Shore Older Adult Mental Health Team, and the Transitional Care Unit at Lion's Gate Hospital. I would also like to acknowledge the work of research assistants Priscilla Ang, Andrea Rahman, and Justine Sanford, who provided assistance in data collection for a portion of the study.

In addition, I would like to thank both the British Columbia Network for Aging Research and The University of British Columbia Faculty of Education for their financial support of the data collection costs of this project provided through a Student Research Data Collection Award and a Graduate Student Research Grant, respectively.

Finally, my deepest gratitude is extended to my parents, Jim and Gloria, my sister, Nicole, and my husband, Pierre, for their support and encouragement of my academic endeavors and for their unending belief in me. A heartfelt thank you is also extended to my dear friends who have witnessed this process unfold and supported me throughout.

Dedicated

to

my parents, Jim and Gloria Myers

Chapter 1 - Overview of Research Study

1.1   Introduction

The measurement and assessment of depression in older adults is a salient issue

due to the debilitating and detrimental affects of depression in the elderly. Despite this,

there are very few measures that have been constructed specifically to screen for

depression in older adults. Quickly administered, easy to use, and accurate screens for

depression in older adults may assist counsellors and health care professionals to

accurately identify those who are suffering from depression, leading to more immediate

diagnosis and treatment for these individuals. Many of the measures that are commonly

used to screen for depression in the elderly have not been designed specifically for use

with this population, even though older adults experience depression and manifest

symptoms in several different ways as compared to younger adults and may experience

physical and cognitive changes that make screening and diagnosis more complex.

Therefore, these measures have some limitations with this population. Care should be

taken when developing depression measures for older adults to not overemphasize

somatic symptoms that might reflect chronic conditions or illness rather than depression,

font size used should acknowledge visual changes that occur with age, and the response

format used should allow for changes in cognitive capacity that might exist for some

older adults.

One measure that has been developed specifically for use as a screening tool for

depression in older adults is the dated and moderately long 30-item measure, the Geriatric

Depression Scale (GDS; Yesavage et al., 1983). There is a current need for an efficient,

up-to-date, and cost-effective screening measure of depression for older adults in clinical

and research settings. It is anticipated that a more recently developed and shorter measure of depression for older adults, the Hubley Depression Scale for Older Adults (HDS-OA; Hubley, 1998), will prove to be at least as useful a measure in screening for depression in this population as the GDS.

The HDS-OA is a quickly administered and short 16-item depression screen designed with older adults in mind that is based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR; American Psychiatric Association [APA], 2000) criteria for depression (A. Hubley, personal communication, July 14, 2008). It uses a dichotomous yes/no response format, large font size, reminders of the reference period, and is freely available for research and clinical purposes. The HDS-OA was designed to be a significant and cost-effective competitor to the GDS and the commonly used, non-population specific depression screen, the Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996). The HDS-OA is a relatively new measure with limited validation evidence and no published validation studies to date. Thus, at present, the HDS-OA requires further validity evidence to support its use.

1.2   Purpose of the Study

This research investigated the construct validity of the HDS-OA. Specifically, this study examined whether the HDS-OA functions (a) as a psychometrically sound and accurate screen for depression and (b) the same or better than the GDS, with a sample of depressed and non-depressed older adults. To answer the above questions, this study examined the reliability of HDS-OA scores, investigated evidence of the convergent and discriminant validity of the HDS-OA through correlation with a measure of a similar

construct and correlations to measures of less related constructs, determined if significant differences existed between depressed and non-depressed group scores, and evaluated ability of the HDS-OA to accurately identify individuals with and without depression. These results were then compared to those for the GDS.

Validation of inferences made from this screening tool will support its future use, which can lead to increased ease and accuracy of screening for depression in older adults by mental health care providers (e.g., psychologists, counsellors, physicians, nurses). The future use of this measure can subsequently improve physical and mental health outcomes by improving the likelihood that older adults will receive diagnostic follow-up as well as timely and appropriate therapy and/or medical attention.

1.3   Overview of Thesis

The following thesis provides a literature review of depression in older adults as well as a review of validation studies of depression measures to guide the methodology of this research. This review is followed by an outline of, and rationale for, the methodology used in the study, including a description of the participants, recruitment, measures used, research design, procedures, proposed data analysis, ethics, and funding. The results of the study are then presented; followed by a discussion of the numerical results, study strengths and limitations, and a discussion of what has been learned about depression and its measurement in older adults. Finally, the thesis concludes with comments on the findings of the numerical analysis, contributions of the research, counselling implications, and areas for future research.

Chapter 2 - Literature Review

2.1   Introduction to the Literature Review

To provide the researcher with an informed background regarding the topic of measurement of depression in older adults, and to guide the development of the research methodology, a literature review was conducted. The presented review begins with a discussion of depression using a broad lens, leading to an overview of depression in older adults. An overview of commonly used depression measures for older adults is then presented, followed by a review of validation studies of depression measures with a variety of populations. Based on the findings in the research literature, the review concludes with an overview of the measure of interest, the HDS-OA, and an introductory rationale for the chosen research design and the selection of other measures used in this study.

Eighteen peer reviewed journal articles addressing depression in older adults and validity studies of depression measures were initially reviewed. This literature was collected through searches of Academic Search Complete database and PsycINFO database search engine using the key terms "depression measures", "depression in older adults", and "depression measure validation". Reviews of these initial articles subsequently led to a review and referencing of further related texts and relevant refereed journal articles published between the years 1960 and 2008.

2.2   Depression in Adults

Depression is the experience of feeling "down" or sad that extends beyond short-lived thoughts of despair or feelings of sadness that last a couple of days into an experience of constant, persistent, and intense negative mood (National Institute of Mental Health [NIMH], 2007). For individuals who report on-going sadness and/or diminished pleasure in activities that they usually enjoy, depression may be suspected (Brody & Serby, 2004). As a result of these experiences, depression interferes with normal functioning and causes pain for the individual suffering from depression, in addition to impacting their close friends and family members (Doris, Ebmeier, & Shajahan, 1999; NIMH, 2007).

Depression is a common illness, with major depression affecting 10-25% of women over their lifetime and 5-12% of men (APA, 2000). Brody and Serby (2004), however, noted that the lower prevalence rate in men may be indicative of a reluctance to report symptoms, leading to the possibility that the prevalence rate is actually higher than the reported rate in men. In addition to its high prevalence, depression is also a serious illness and is one of the leading causes of the loss of years of healthy life, second only to cardiovascular disease (NIMH, 2007). In fact, the American Psychiatric Association (2000) stated that up to 15% of individuals with severe major depressive disorder die by suicide. NIMH (2007) stated that a large proportion of those with a depressive disorder do not seek treatment, but the majority of individuals with depression can get better with treatment, even those with severe depression. Research into the disorder has resulted in the development of medications, psychotherapies, and other methods to treat people suffering from depression (Doris, Ebmeier, & Shajahan,

1999). The definitive cause of depression is unclear, but it is becoming evident that a number of factors may be responsible, including genetic (e.g., family history of depression), neurobiological (e.g., dysregulation in neurotransmitter systems or hormonal imbalances), and environmental factors (e.g., stressful life events) (APA, 2000; Doris et al., 1999).

Diagnosis of major depressive disorder (MDD) or another form of depressive disorder depends on the severity, chronicity, and persistence of sadness and/or loss of pleasure and the number of additional symptoms that are present (APA, 2000; First et al., 2007). Consequently, there are several forms of depressive disorders. The most common are MDD and dysthymic disorder (Doris et al., 1999; NIMH, 2007).

MDD is a clinical course that is characterized by a single or recurrent major depressive episode (MDE) (APA, 2000). MDE is defined by symptoms that interfere with an individual's ability to work, sleep, study, eat, and enjoy activities (Brody & Serby, 2004; NIMH, 2007). MDE also prevents a person from functioning normally in the social and occupational realms (APA, 2000; NIMH, 2007; Brody & Serby, 2004). As well, MDE may occur only once in a person's lifetime but, more often, it recurs throughout a person's life (NIMH, 2007). As identified by the DSM-IV-TR, the diagnostic criteria for a current major depressive episode include depressed mood and/or a loss of interest in activities that are usually enjoyed, for most of the day, nearly every day, for a period of at least two weeks over the past month (APA, 2000). If only one of these criteria are present, it must be accompanied by at least four of the following symptoms during the same two week period: significant weight loss or weight gain (of at least 2.5%) related to appetite changes, a change in sleeping patterns

leading to insomnia or hypersomnia, psychomotor retardation or agitation that is observable by others, fatigue or loss of energy, feelings of worthlessness or inappropriate or excessive guilt, reduced ability to think or concentrate or make decisions, and/or recurrent thoughts of death or suicidal ideation or suicide plan or attempt (APA, 2000). These additional symptoms must also persist for most of the day, nearly every day, for the two-week period. If depressed mood and loss of interest in pleasurable activities are both present, at least three additional symptoms are required.

Compared to a MDE, dysthymic disorder, also known as dysthymia, presents with less severe symptoms that may not be disabling but they can affect normal functioning and reduce feelings of well-being (Doris et al., 1999). For dysthymia to be diagnosed, a chronically depressed mood must be present for a period of two years or more, on more days than not, accompanied by at least two of the following symptoms: poor appetite or overeating, insomnia or hypersomnia, low energy or fatigue, low self-esteem, difficulty concentrating or making decisions, and feelings of hopelessness (APA, 2000; Brody & Serby, 2004). People with dysthymia may also experience one or more episodes of major depression during their lifetime (NIMH, 2007). A MDE can also be superimposed on dysthymia (Doris et al., 1999).

A further depressive disorder category includes depressive disorder not otherwise specifed (NOS) and includes depressive disorders that do not meet the criteria for MDD or dsythymia. These include, for example, premenstrual dysphoric disorder, minor depressive disorder where depressive symptoms last for at least two weeks but fewer than five symptoms are present, recurrent brief depressive disorder

where symptoms last two days up to two weeks and occur once a month for 12 months, or postpsychotic depressive disorder of schizophrenia (APA, 2000).

Some forms of depressive disorder present with different characteristics than those already described, or they may develop under special circumstances (APA, 2000; NIMH, 2007). These forms include a psychotic depression, which occurs when a severe depressive disorder is superimposed on a form of psychosis, such as disassociation or the experience of hallucinations or delusions (American Psychiatric Association, 2000; NIMH, 2007). Another form is postpartum depression, which may be diagnosed if a new mother develops a MDE within one month after delivery. The prevalence rate for post-partum depression is estimated to be as high as 10% to 15% (NIMH, 2007). An additional form of depressive disorder is seasonal affective disorder (SAD), which is defined as a depressive disorder that begins during the winter months and subsides when seasons change and daylight increases (American Psychiatric Association, 2000; NIMH, 2007).

Finally, bipolar disorder, also called manic-depressive disorder, is not as common as MDD or dysthymia, but is characterized by moods that cyclically swing between extreme highs (mania) and extreme lows (depression) (Brody & Serby, 2004; NIMH, 2007).

2.3   Overview of Depression in Older Adults

Sable, Dunn, and Zisook (2002) reported that depression is the most common mental illness among individuals over the age of 60 and approximately 15% of older adults suffer from a significant depression in their lifetime. Women are more likely to

be affected than men at older ages (Djernes, 2006; Vink, Aartsen, & Schoevers, 2007), but this difference narrows with advancing age (Gatz & Fiske, 2003; Teper & Thomas, 2006). Although the rate of MDE is generally lower among older adults when compared with younger adults, prevalence rates of depressive symptoms are 20-30% and the rates for those with chronic physical illness are estimated to be 70% (Blazer, Burchett, Service, & George, 1991). Jeste, Alexopoulos, Bartels, and Cummings (2008) concur with this prevalence rate for depressive symptomatology and stated that, although the rate of diagnosable depression in older adults is relatively low, the prevalence of clinically significant levels of depressive symptoms is high at approximately 15-20% of all older adults. A review of studies of prevalence rates of depression in older adults revealed an overall prevalence rate of 1% to 9.4% in community households and 14 % to 42% in institutional settings (Djernes, 2006).

It has also been reported that 18% of individuals in the United States who are 65 years and older commit suicide, as compared to a 13% rate of suicide in the general population (NIMH, 2003). Caucasian men 85 years and older have the highest rate of suicide. Boyd and Bee (2006) reported that individuals 65 years and older tend to be vulnerable to chronic or episodic depression. In addition, results of a study examining the effects of mattering to others, purpose in life, and depression on overall wellness in 167 older adults living in the Southwestern U.S. showed that these variables accounted for just over 78% of the variance in overall wellness for older adults (Dixon, 2007). Given these findings, it appears that depression is a significant issue in older adults.

In fact, depression in later life is associated with mortality due to life losses, such as the loss of employment, sexual enjoyment, and cognitive functioning (Pitt,

1998). Depression also compounds the negative effects of physical illness that are often experienced by older adults. Callahan, Hui, Nienaber, Musick, and Tierney (1994) reported that, when compared to non-depressed older adults with similar physical illness, depressed older adults have increased rates of emergency room services for psychiatric and non-psychiatric issues (40% versus 29% for non-depressed individuals), 38% more outpatient visits, and 61% higher outpatient costs. Overall health-care costs of depressed older adults are double those of non-depressed older adults with similar illness (Simon, von Korff, & Barlow, 1995).

With respect to etiology, many studies agree that depression in older adults may develop as a result of increased stress due to bereavement, relocation, care giving, cognitive and physical impairment, interpersonal problems, low social support, and/or medical conditions (Gatz, 2000; Gatz & Fiske, 2003; Kraaij & Garnefski, 2002; Vink et al., 2007). Djernes (2006) concurred and added that a history of depression is also a predictor of depression in the elderly and Katz (2000) suggested that familial history of depression was especially prevalent in early-onset depression. Vink et al. indicated that personality traits, dysfunctional functioning, and negative self-image play a role in the development of depression in older adults, and Gatz and Fiske implicated genetic factors. It has also been reported that biological factors play an important role in the etiology of depression in the elderly; for example, there is a link between vascular pathology, neurological illness, and depression (American Psychiatric Association, 2000; Teper & Thomas, 2006). Katz agreed and noted the prevalence of white matter lesions in the brains of older adults with late-onset depression.

As noted, medical conditions were found to act as a strong predictor of depression in the elderly. In fact, Serby and Yu (2003) identified that 40% of adults with Parkinson's disease, 20-25% of stroke victims, 15-20% of those with Alzheimer's disease, and 50% of those with dementia have significant levels of depression. Other co-morbid medical conditions include cancer, cardiovascular disease, metabolic disturbances, arthritis, and sensory loss (NIMH, 2003). Teper and Thomas (2006) indicated that prognosis is poor if co-morbid depression is not identified and treated, and may lead to a higher risk of dementia and death.

Although depressive symptomatology occurs with similar frequency in older adults as compared to younger adults, there are some distinct differences that differentiate the experience of depression in the elderly from depression that occurs at younger ages. These differences include a development period for depression of potentially years in older adults versus days or weeks in younger adults (Berger, Small, Forsell, Winblad, & Blackman, 1998; Huffstetler, 2001), increased loneliness experienced as a result of the loss of family members and friends, and a reduction in physical functioning that compounds the effects of depression (Serby & Yu, 2003). As well, a lack of diagnosis and treatment of depression in older adults can lengthen and increase the severity of the depressive illness (Gatz, 2000; Serby & Yu, 2003).

Many of the symptoms of depression in older adults are similar to those seen in younger adults (APA, 2000), but research has demonstrated some significant differences in symptomatology between these groups. Brodaty, Peters, and Boyce (1991) found that older adults were more severely depressed and had more delusions, psychomotor disturbances, agitation, and appetite loss. A subsequent study showed that

older adults also had more somatic symptoms and guilt (Brodaty, Luscombe, & Parker, 1997). The dysphoric symptoms of depression may also present differently in older adults. Rather than describing that one is feeling "down", an older adult may express pessimism or helplessness, or show signs of apathy or self-deprecation (Cavanaugh & Blanchard-Fields, 2002). Katz (2000) noted that the most consistent affective symptom of major depression in older adults might actually be anhedonia (i.e., the absence of positive affect) rather than an increase in negative affect. Kraaij and Garnefski (2002) also found that depression in older adults might also manifest itself behaviorally through insecure attachments to loved ones. As well, impairments in attention, memory, and executive function have also been confirmed to be more prevalent in older adults with depression (APA, 2000; Lockwood, Alexopoulos, & Van Gorp, 2002). In addition, a study on depression in later life showed that depressed older adults also experience more symptoms of anxiety (Baldwin & Tomenson, 1995).

Stickle and Ondera (2006) recommended that, in addition to screening for the depressive symptoms specifically seen in older adults, a more comprehensive assessment should also be conducted into how the distress is expressed and the impact of physical illness on depressive symptoms. The authors also noted that medication usage, medical conditions, and issues of substance abuse should be identified. In assessing depression in older adults, it is also important to investigate excessive complaints of pain that may be masking symptoms of depression and to not simply attribute a loss of sex drive to the aging process, social isolation to the passing away of friends with age, or extended grief to the passing away of a spouse rather than attribute these symptoms to depression itself (Stickle & Ondera, 2006).

12

Due to the slow development of depression in the elderly and the manifestation of the symptoms of depression often as physical signs rather than overt sadness, the disorder is often not detected (Huffstetler, 2001). The co-morbidity of medical conditions with depression also makes depression screening a challenge (Stickle & Ondera, 2006). The ability to screen for depression begins with more attention being given to identifying the signs and symptoms of depression in older adults and the use of more efficient screening measures that are designed specifically to identify depression in this population. In addition to reducing health care costs, an efficient screen may allow for the recognition and subsequent treatment of depression at an earlier stage, which may lead to an increase in wellbeing and a reduction in suicide rates in older adults. This prediction is made based on the findings of Uncapher, Gallagher-Thomas, Osgood, and Bongar (1998), who reported that, in older adults, depression is more predictive than hopelessness of suicide.

2.4   Measures of Depression in Older Adults

Preceding a review of select validation studies conducted for commonly used depression measures, this section will review the diagnostic tools and measures most commonly used to screen for depression in older adults, with the intent of providing an overview of the measures currently available for use with this population.

The majority of depression screens were not developed specifically for use with an older population, despite the differences in manifestation of the symptoms of depression and the experience of depression in older adults. Further issues with the use of generic depression screens with older adults include variability in cognitive functioning,

decision-making, and visual acuity in this age group as compared to younger adults, and increases in the presence of co-morbid medical conditions and physical impairment in older adults that can complicate the process of screening for depression in this population. Only one currently used self-rated depression measure appears to have been designed specifically to screen for depressive symptomatology in older adults. This measure, the GDS, will be discussed and an overview of its limitations will be explored, with the intent of highlighting the need for the validation of a new measure of depression for older adults, thus supporting the purpose of this study. As a result of the limited array of population specific measures from which to choose, non-population specific depression measures are often utilized to screen for depression in older adults. An overview of these measures will be provided, as well as a description of the context of their use with older adults and their inherent limitations. Finally, the HDS-OA, a relatively new measure that has been designed specifically for use as a self-rated depression screen for older adults will be introduced.

2.4.1   Diagnostic measures

The DSM-IV-TR suggests that the diagnosis of depression in older adults involves the same criteria as it does for any age group but cautions that, in older adults, these symptoms can vary and may be complicated by the aging process (APA, 2000). A review of the literature showed that the Structured Clinical Interview for DSM-IV-TR Axis-I Disorders, Non-patient Version (SCID-I/NP; First et al., 2007), or an older version of this interview, is often used as the "gold standard" to diagnose depression in a non-patient population. The SCID-I/NP is a semi-structured interview based on DSM-IV-TR

criteria and it was designed to assist researchers and clinicians in making Axis-I mood disorder diagnoses (First, Gibbon, Spitzer, & Williams, n.d.).

The SCID-I/NP was used as the criterion measure in several depression measure validation studies with older adults, including Low and Hubley's (2007) validation study of two depression screens administered to 119 older cardiac patients. An older version of the SCID was used in Olin, Schneider, Eaton, Zemansky, and Pollock's (1992) validation study of a common depression measure administered to 25 older adult outpatients with major depression and 25 healthy control subjects. In a similar vein, Lewinsohn, Seeley, Roberts, and Allen's (1997) study of a depression screen for older adults utilized DSM-III-R criteria for depression, the depression criteria on which the SCID-I/NP is based, to make a diagnosis of depression or confirm the absence of depression in each participant.

2.4.2 Depression screens designed for use with older adults

To date, there is only one commonly used self-rated depression measure that appears to have been designed specifically for use with older adults. This measure is the 30-item GDS. The GDS is a self-administered measure of depressive symptoms, composed of items that focus on affective and cognitive symptoms that have been experienced over the past week. The measure uses a yes-no response format and includes 10 reverse-scored items (Yesavage et al., 1983). The GDS has been shown to be a useful screening tool for depression in older adults due to its high specificity and sensitivity (Katona & Livingston, 2000).

Yesavage et al. (1983), in a report on the development and validation of the GDS with a sample of 40 non-depressed elderly community participants and 60 elderly

subjects being treated for depression, noted that higher than usual somatic complaints as well as cognitive complaints in this population present an opportunity and a hindrance in screening for depression. The authors noted that most depression scales contain a high proportion of items that measure somatic symptoms. When screening for depression in older adults, the authors discovered that it might be necessary to weigh somatic symptoms of depression less heavily than psychological symptoms. As well, the authors recognized that the unique cognitive complaints of older adults may provide the chance to develop a screening tool with a greater ability to discriminate the depressed from the non-depressed in elderly populations. This knowledge led to the development of the GDS as a measure of depression in older adults that focuses on affect and cognitions rather than somatic complaints (Yesavage et al., 1983). As a result of this method of development, the measure is not based on DSM diagnostic criteria.

In comparing the GDS to the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, & Erbauch, 1961), Olin et al. (1992), in a study of 25 older adult outpatients with major depression and 25 healthy control subjects, reported that both the BDI and the GDS appeared to be efficient self-report instruments in screening for depression in an outpatient clinic. Despite the reported screening efficiency of both measures (i.e., the ability of the measure to accurately detect the presence or absence of the disorder in the population that is being screened (Lewinsohn et al., 1997)), the authors noted that depressed participants more frequently selected multiple responses on items of the BDI, reflecting the increased difficulty that older adults may have in making decisions when using a multiple-choice response format and suggesting that the yes-no response format of the GDS items is easier for older adults to answer.

Similarly, Low and Hubley (2007) compared the BDI-II and the GDS as screening measures for depression in 119 older cardiac patients. The results demonstrated that both measures possessed excellent sensitivity for detecting both major depressive disorder and double depression in a cardiac sample, but the GDS had greater positive predictive values and higher specificity, and thus was able to better differentiate those who were depressed from those who were not depressed. These results support the use of the GDS over the BDI-II in an older population with prevalent somatic symptoms.

The GDS has also been compared to the Zung Self-rating Depression Scale (SDS; Zung, 1965). Yesavage et al. (1983), in their discussion of the development and initial validation of the GDS, reported an internal consistency of .94 for the GDS as compared to .87 for the SDS and a better split-half reliability of .94 for the GDS as compared to .81 for the SDS. The authors also reported that the SDS discriminated less effectively between normal, mildly depressed, and severely depressed individuals when compared to the GDS. Moreover, in a subsequent study that compared the SDS to the GDS in a sample of 439 community-dwelling older adults, Dunn and Sacco (1989) concluded that the GDS was superior to the SDS in terms of non-completion rate and convergent validity with the Depression Symptom Checklist (DSC; Sacco, 1983), indicating that the yes-no format of the GDS may be easier for older adults to use than the ordered response format of the SDS.

2.4.3   Non-population specific depression screens used with older adults

A review of the literature also revealed that several non-diagnostic and general population-based screening measures for depression have often been used in the context

of depression research with older adults. In particular, three screening measures (the 21-item, multiple-choice BDI and subsequently developed BDI-II; the 20-item, ordered-response format SDS; and the 20-item, ordered response format Center for Epidemiological Studies Depression Scale [CES-D; Radloff, 1977]) have been demonstrated to be useful in identifying older adults who are at-risk for depression, despite not being designed specifically for use with this population. Although these measures show utility, there are limitations to their use with older adults. An overview of each measure, as well as their strengths and weaknesses, is presented below.

Beck Depression Inventory. In a review of the development of the BDI and an examination of the utility of this measure with older adults, Gallagher (1986) explained that, during its development, items were clinically derived and were chosen as a result of their ability to discriminate depressed from non-depressed patients and provide an accurate gauge of symptom severity. Gallagher indicated that the final measure consisted of 21 categories of symptoms, each describing a cognitive, behavioral, or affective manifestation of a depressive symptom that may have been experienced in the past two weeks, with four response option statements of the symptom ranging from 0 (not present) to 3 (intensely present). The total score can range from 0 to 63. Although numerous validation studies of the BDI provide evidence of reliability and validity for clinical and research purposes with general adult populations, Gallagher raised concerns over the inclusion of the number of somatic items in the BDI, which may lead to inflated depression scores in a non-depressed, yet physically ill, population. The review also highlighted concerns with the readability of the BDI when used with an older population, who may have less formal education or possess mild cognitive impairment. The BDI was

later revised and the number of somatic items was reduced. A subsequent study of the psychometric properties of the revised BDI-II among 376 community-dwelling adults, with a focus on a subgroup of 147 older adults aged 55-90, revealed that the updated BDI-II appeared to have strong reliability and validity evidence supporting its use as a screening measure for depression in older adults (Segal, Coolidge, Cahill, & Riley, 2008).

Center for Epidemiological Studies Depression Scale. As previously highlighted, some of the concerns regarding the use of depression screens with older adults involve the use of these measures with individuals who have a concurrent disease or a cognitive or functional impairment, in addition to concerns over the use of a measure with individuals who may respond to items in a socially desirable manner due to social stigma or cultural background (Lewinsohn et al., 1997). Any of these aspects can bias results. Despite these concerns, the CES-D has demonstrated efficacy with older adults under these circumstances, even though it was designed for use with the general population. The CES-D is composed of 20 items that are designed to measure current levels of depressive symptomatology experienced in the past two weeks, with a focus on the affective symptoms of depression (Radloff, 1977). Each item is scored from 0 to 3, based on frequency of the occurrence of each symptom, with 3 being the highest occurrence. The resulting total score, therefore, can range from 0 to 60. Radloff reported that the scale possesses high internal consistency, good test-retest reliability, and adequate validity. A study of the efficacy of the CES-D as a screen for clinical depression with 1005 community dwelling older adults, aged 50 to 96, showed that none of age, gender, cognitive impairment, functional impairment, physical disease, or social disability had an adverse effect on the psychometric properties or screening efficiency of the CES-D

(Lewinsohn et al., 1997). Although these results are promising, research on this scale is limited with older adults and the authors also suggested that further investigation be conducted into similar measures that screen for depression in older adults.

The Zung Self-rating Depression Scale. The SDS is another measure that is frequently used to screen for depression in the elderly, despite being developed for use with younger adults (Dunn & Sacco, 1989). This scale is composed of 20 items, with four response choices assessing the frequency of occurrence of symptoms that include "none", "a little of the time", "most of the time", and "all of the time" (Dunn & Sacco, 1989; Zung, 1965). In a study comparing the SDS to the GDS, Dunn and Sacco (1989) reported a good internal consistency of .84 for the SDS when used with a sample of 439 community-dwelling older adults, with a mean age of 74 years. Despite good internal consistency, a lower value was found for its concurrent validity as compared to that for the GDS, and it had a lower item completion rate. The authors hypothesized that the response format of the SDS may be more difficult for older adults to use as compared to the easy to use yes-no format of the GDS.

Although the BDI-II, the CES-D, and the SDS show some good reliability and validity evidence for use as depression screens with older adults, this evidence is limited and they have not been designed for use with this population. As a result, these measures possess the noted limitations.

2.4.4   A new depression screen designed for use with older adults

A relatively new measure, the HDS-OA, is a 16-item, self-administered screen for depression in adults aged 55 and older. The HDS-OA is also comprised of two

additional (non-scored) questions used to capture information about current

bereavement and new medication use. HDS-OA items are based on DSM-IV-TR

criteria for major depressive disorder and dysthymia and therefore contain some

somatic items, but a yes-no response format is used for all questions, it uses a large font

size, and each question contains a reminder of the previous two-week time frame within

which to frame the answer (A. Hubley, personal communication, July 14, 2008). Total

scores can range from 0-16 and the measure is easily scored using an answer key, with

responses that are in concurrence with depressive symptomatology each receiving a

score of "1".

The HDS-OA has been the subject of two validation studies to date. The HDS-

OA was administered in a recent study that examined the psychometric properties of

several measures of depression with a sample of 119 cardiac patients. The manuscript

for this study is under development. In addition, another study has recently been

completed that examined the psychometric properties of the BDI-II and the HDS-OA

with a sample of 50 depressed and 50 non-depressed middle aged and older adults.

Preliminary results indicated high internal consistency, high convergent validity with

the BDI-II, a statistically significant difference in scores between depressed and non-

depressed groups, and high sensitivity and specificity (Hubley, Mangaoang, Burke, Ho,

Ang, Myers, & Chiu, 2009). Despite these positive findings, further validation evidence

is required to support the use of the HDS-OA as a screen for depression in older adults.

2.4.5   Summary

As the research literature demonstrates, only the GDS has been designed specifically as a screening tool for depression in older adults. Its focus on affective and cognitive symptoms alleviates the concerns researchers have had about the inclusion of a high number of somatic items often seen in general population-based depression screens, which may confound results when used with older adults. As well, the GDS is an easily administered and easy to use measure due to its yes-no format and clear language. Despite these advantages, the GDS has its limitations, including the 30-item length of the original version and, more importantly, its lack of association with the DSM-IV two-week time frame and criteria for depression.

There are several other measures that have been used efficiently as screens for depression with older populations. These measures include the BDI, BDI-II, CES-D, and SDS, although they were not specifically designed for use with this population and therefore also possess limitations. These drawbacks include limited research with the population of interest in the case of the CES-D and SDS, concerns over the response format and completion rate of the SDS, and concerns over the use of somatic items, readability, multiple choice item format, and completion rate of the BDI-II.

The lack of depression screen alternatives that are specifically designed for use with an older sample, the limitations associated with the GDS, and the difficulties described for other depression screens not designed for use with an older sample all support the intent of the study, which is to examine validity evidence for the HDS-OA, a relatively new measure of depression designed to be used with older adults.

2.5   Review of Depression Measure Validation Studies

This section will review research literature that further compares and provides validation evidence for commonly used depression measures across a variety of populations, with the intent of highlighting the most sound and often used approaches to validation that will subsequently guide the approaches used in this study to provide further validation evidence for HDS-OA. This review will also assist in guiding and supporting the rationale of the selection of specific procedures used in this study.

The validation articles selected for review were chosen based on inclusion of the most commonly used and significant depression measures and cover the last 25 years (i.e., 1983 to 2008). Validation studies previous to this time period were not included for review due to the changes that have occurred since the early-1980s with respect to developments in validation methods, which are, in part, due to updates in statistical methodology and a gradual shift in the view of validation from a trinitarian to a unitary concept. These concepts are described in the following section.

2.5.1   Approach to the review of validation studies

The trinitarian view of validity, which evaluates content, criterion, and construct validity separately, has dominated psychology for almost a half century. It outlines that there are many types of validation, with each one often presented as being solely adequate to support validity depending on the nature of the measure or its application (Hubley & Zumbo, 1996; Shultz, Riggs, & Kottke, 1998). In contrast, in this review and in this research, a modern approach is being embraced. This approach is the unitary concept of validity, which appeared in the late 1970s and early 1980s and has since been

gaining support. This concept outlines that validity and construct validity are one and the same, with validity as a single construct that may be supported by many different lines of evidence (Hubley & Zumbo, 1996; Schultz et al., 1998). In particular, it describes that the accumulation of a body of validation evidence for a measure is required before a judgment can be made about its validity for use with a particular population, with each piece of evidence lending to the process of validation. This is in contrast to the trinitarian view, where one type of validation evidence supports the validity of a measure. The unified concept of validity incorporates other types of validity, such as content and criterion validity into a construct-based framework to test hypotheses about score meaning and suspected relationships (Messick, 1995). Construct validity can be defined as the appropriateness of inferences made about a latent quality or characteristic that differs in individuals (Gregory, 2004; Messick, 1995). Included under construct validity are the approaches of factorial, convergent, and discriminant validity, as well as reliability.

Despite the chosen approach for this review, it was identified that a majority of the reviewed depression measure validation studies have taken a trinitarian-like approach to the presentation of validation evidence. Many have presented content, criterion, and construct (i.e., factorial, convergent, and discriminant validity evidence) in separate studies for each measure of focus, but some studies have combined several validity approaches into one study to build a body of validation evidence for a given measure. In this review, a variety of validation procedures will be examined for depression measures that have frequently appeared in the research literature. Given the compartmentalized presentation of validation evidence in the reviewed studies and ease of discussing the

validation evidence in this manner, the validation procedures reviewed will be categorized and examined according to the type of validation evidence provided. Consistent with the unitarian approach to validity, however, it is noted that all of the validation evidence examined falls under the concept of construct validity. When these multiple pieces of validation evidence are combined for any particular measure used with a specific population, it provides for a fuller picture and stronger support of the validity of inferences from the measure with that population.

2.5.2   Content validity evidence

Content validity evidence demonstrates that a measure provides a good sample of the behavior that it is intended to measure, and it is determined by the degree to which the questions or items on a test encompass these behaviors (Gregory, 2004). In a similar vein, Haynes et al. (1995) define content validity as the degree to which elements (e.g., items, response formats, instructions) of a measure are relevant to and representative of the full scope of the construct that is being measured. In this light, content validity has an impact on the validity of inferences that can be drawn from the collected data and, as per the unitary concept of validity, is an element of construct validity.

In a review of depression measure validation studies, validation studies providing evidence of content validation were limited. In fact, a search of the PsycINFO database using the keywords "content validity" and "depression" revealed one depression measure validation study. Four studies in total were identified for examination upon a further detailed search.

Osman, Kopper, Barrios, Gutierrez, and Bagge (2004) provided content validity evidence for the BDI-II. The DSM-IV was used as the evaluative standard by expert raters, since BDI-II items were developed to identify depressive symptoms based on the DSM-IV criteria for depressive disorders. The number of participants in this study included seven expert raters as well as 13 DSM-IV Axis-I and -II diagnosed adolescent psychiatric inpatients who evaluated the ease of reading, understanding, and usefulness of each item in identifying depression. Osman et al. (2004) gathered expert evaluations of the appropriateness of items with respect to DSM-IV criteria and their specificity to major depressive disorder symptoms. They also gathered target population-based evaluations of the clarity and usefulness of items on a 5-point rating scale. They did not, however, subject all the elements of the instrument to content validation. For example, the instructions and the response format were not evaluated. The authors did examine the proportional representation of items though, with results suggesting the inclusion of more behavioral items and the need to drop items with low relevancy that did not correspond directly to the DSM-IV symptoms of major depressive disorder, such as past failure, punishment feelings, and loss of interest in sex.

Similarly, Onega (2008) provided content validity evidence for the Depressive Symptom Assessment for Older Adults (DSA; Onega, 2006). The DSA is 27-item, interviewer-rated screening instrument, designed in 1998 to measure depressive symptoms in older adults, regardless of their cognitive status (Onega, 2008). Content validity evidence was gathered using a two-phased Delphi technique, with the measure being revised after each phase, based on percent agreement. In phase one, seven non-depressed seniors from a retirement community served as content validation experts.

Once changes were made to the DSA, the same experts again served as content validation experts for phase-two. In contrast to Osman et al. (2004), Onega (2008) did not utilize a second group of expert raters who were professionals in the field of depression or psychometrics, but rather used a community sample as lay experts. Onega (2008) justified the inclusion of lay experts by stating that, by interviewing older adults who acted as content experts, problems with how depression was measured in this population could be determined. In addition, it did not appear that the author evaluated the instructions for the DSA as part of content validity and it was not clear if a formal scaling procedure to quantify evaluations was utilized. It was evident, though, that the author did use standard content validation questionnaires for both phases and questionnaire items were designed to obtain evaluations of the clarity and internal consistency of the DSA. The questionnaire also provided definitions of the aspects of content validity that were being evaluated and definitions of each subscale for the DSA. Onega (2008) also examined the proportional representation of items, which resulted in the addition of another scale, and indicated that further psychometric testing was underway to provide further construct validation evidence, in accordance with content validation guidelines discussed in Haynes et al. (1995).

Two further studies were reviewed that provide a brief overview of content validation procedures and a degree of content validation evidence for two depression measures. Yesavage et al. (1983) discussed the process of initial item selection for the GDS, which involved an expert team of geriatric psychiatry clinicians and researchers who generated initial items for the measure based on a wide variety of topics related to depression. The scale was then administered to 47 community subjects and the items that

best correlated with the total score were chosen for inclusion in the GDS. The number of expert raters was not identified. Similarly, Nyenhuis, Luchetta, Yamamoto, and Terrien (1998) discussed the content validation process for the Chicago Multiscale Depression Inventory (CMDI; Nyenhuis et al., 1998), a self-report depression measure used to differentiate between mood, self-evaluative, and vegetative components of depression. Content validation evidence for this measure was provided by the evaluation of the author's initial list of 87 items by 10 doctoral level licensed psychologists, who classified each item into one of the mood, self-evaluative, and vegetative categories of depression. Where agreement was less than 80%, items were discarded, resulting in 75 items that were administered to depressed and non-depressed groups. Items were then retained if high correlations were obtained with the sub-scale total and with the BDI, resulting in a total of 42 items. In both of these studies, a panel of expert raters was used to either select items or evaluate the relevance and appropriateness of each item for the measure.

In the majority of the reviewed studies, the number of content review experts was sufficient as between three and ten content review experts are recommended to participate in content validation studies (Lynn, 1986). In addition, members of the target population were also used as subject matter experts in several studies to increase the chance that the items were relevant to the construct. As well, the examination of the proportional representation of items, one of the guidelines in the approach to providing content validation evidence, was incorporated into the majority of these studies to ensure that the items were distributed in a way that reflected the different facets of the construct under consideration (Haynes et al., 1995). It was evident, though, that several important components of content validation were omitted from some of the studies, including the

content validation of all elements of a measure and the use of a formalized scaling procedure for evaluation. For studies that were not conducted during the initial development phase of the measure, it cannot be determined if other important aspects of content validation were addressed, such as defining the domain of the construct before developing other elements of the measure and the use of population and content experts for initial generation of items and other measure elements (Haynes et al., 1995).

2.5.3   Criterion validity evidence

Criterion-related validity evidence is provided when a measure is shown to be effective in estimating performance on an outcome measure (Gregory, 2004). In concurrent validity, criterion measures are obtained at the same time as the test scores. Predictive validity, in contrast, obtains the criterion measures at some time in the future. Several studies that were reviewed provided validation evidence for common depression measures using a criterion validation approach. These include studies that focused on the CES-D (Lewinsohn et al., 1997), the GDS (Dunn & Sacco, 1989; Low & Hubley, 2007; Olin et al., 1992; Yesavage et al., 1983), the GDS-5 (Marquez, McAuley, Motl, & Elavsky, 2006), the BDI (Olin et al., 1992), the BDI-II (Low & Hubley, 2007), and the SDS (Dunn & Sacco, 1989).

In several studies (Dunn & Sacco, 1989; Lewinsohn et al., 1997; Low & Hubley, 2007; Marquez et al., 2006; Olin et al., 1992), concurrent validity evidence was provided through an evaluation of screening efficiencies by the calculation of sensitivity, specificity, positive predictive values (PPV), and negative predictive values (NPV). Based on the results of a criterion measure, for any given cut score, sensitivity is the

probability of a true positive outcome on a measure of interest, specificity is the probability of a true negative outcome on a measure of interest, PPV is the proportion of positive outcomes of a measure that are identified as positive by the criterion measure, and NPV is the proportion of negative outcomes of a measure that are identified as negative by the criterion measure. Wassertheil-Smoller (1990) indicated that, when a disorder has a low prevalence in the population, the PPV is also low. Therefore, in evaluating or determining a cut score for a measure, an appropriate cut score should be selected that provides a balance between high sensitivity, specificity, and NPV and as high a value for the PPV as possible and one that does not result in the lowering of sensitivity and specificity values considerably (Swets et al., 2000). The selection of an optimal cut score for a screen will depend on the clinical context in which it is used (Hsiao, Bartko, & Potter, 1989). In the case of determining an appropriate cut score for a depression screen, more weight has often been placed on sensitivity versus specificity to ensure that truly depressed individuals are accurately identified (Almeida & Almeida, 1999; Low & Hubley, 2007). See Figure 2.1 for a matrix describing these values.

Figure 2.1 Description of Sensitivity, specificity, PPV, and NPV.

		Criterion Measure Outcome		
		Depressed	Non-depressed	
Screening Measure Outcome	Depressed	True Positive (TP)	False Positive (FP)	PPV = TP/(TP+FP)
	Non-depressed	False Negative (FN)	True Negative (TN)	NPV = TN/(TN+FN)
		Sensitivity = TP/(TP+FN)	Specificity = TN/(FP+TN)	

In several studies (Lewinsohn et al., 1997; Low & Hubley, 2007), it was stated that these values were obtained via receiver operating characteristic (ROC) curve analyses. The ROC curve plots the proportion of true positives against the proportion of false positives for each identified cut-score on the measure of interest (Swets, Dawes, & Monahan, 2000). The resulting area under the curve (AUC) is an indicator of the accuracy of a diagnostic scale and its value can range between 0.0 and 1.0 (Hsiao et al., 1989). An AUC of .80 or higher demonstrates the usefulness of a measure as a diagnostic screen (Holmes, 1998). The remainder of the studies provided many of these values, although they did not explicitly state that an ROC curve analysis was performed, but rather referred to the analysis as an examination of the sensitivity and specificity of a measure or a determination of the screening efficiency of a measure.

Criterion validity evidence has also been provided through examination of group differences based on diagnostic categories (Low & Hubley, 2007; Olin, et al., 1992; Snyder, Stanley, Novy, & Averill, 2000; Yesavage et al. 1983). For these studies, differences in mean scores between categories of depressed and non-depressed groups were examined using an appropriate mean score difference test, such as a between-groups t-tests (Low & Hubley, 2007; Snyder et al., 2000) or a between groups analysis of variance (ANOVA) (Olin, et al., 1992). In Yesavage et al., where the mean scores of more than two groups were being compared at one time, an F-test was first utilized to detect a statistically significant difference in mean scores between any of the groups, followed by a post-hoc t-test to detect statistically significant differences between each group.

The investigations into screening efficiencies and group differences were made possible through the use of diagnostic interviews that served as a criterion measure, such as the Research Diagnostic Criteria (RDC; Spitzer, Endicott, & Robins, 1978), which has been used to confirm the absence or presence of depression in clinical and community samples in several studies (Lewinsohn et al., 1997; Yesavage et al. 1983). The Structured Clinical Interview for DSM-III-R, Patient Edition (SCID-P; Spitzer, Williams, & Gibbons, 1986) has also been used by as a criterion measure to confirm the presence of depression in a clinically depressed sample by Olin et al. (1992). The authors also used the non-patient edition of the Structured Clinical Interview for DSM-III-R (SCID-NP; Spitzer, Williams, & Gibbons, 1988) to confirm the absence of depression in a community sample. The more updated version of the Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID-I/NP; First et al., 2002) was also used in several studies (Low & Hubley, 2007; Hubley et al., 2009) as a criterion measure to confirm the absence or presence of depression in patients and community samples.

In contrast, in providing validation evidence for the GDS-5 with a sample of 185 sedentary older adults, the criterion measure used by Marquez et al. (2006) was the 30-item GDS, rather than a "gold standard" measure such as a diagnostic interview. The use of the GDS as the criterion measure may lend less credibility to the results of this particular study because it is a self-report instrument that is still accumulating validation evidence, not a diagnostic interview that possesses strong validation evidence supporting the inferences made from the results. The GDS also has item overlap with the GDS-5, so criterion contamination will also be an issue. Similarly, Dunn and Sacco (1989) reported concurrent validity evidence for the GDS and SDS by comparing scores of these

measures with the Depression Symptom Checklist (DSC; Sacco, 1983) approximation of the DSM-II criteria for a major depressive episode, rather than using a diagnostic interview as the criterion measure. The authors indicated that criterion validity evidence has been reported for the DSC, but the authors did not provide these validation study references.

Several studies used both a clinical and a control group in their validation studies, aiding in the determination of group differences. Olin et al. (1992) utilized a sample of 25 depressed older adult outpatients and 25 non-depressed community dwelling older adults. Similarly, Yesavage et al. (1983) utilized a sample of 40 non-depressed older adults and 60 depressed older adults in their study. Although neither Dunn and Sacco (1989), Lewinsohn et al. (1997), nor Low and Hubley (2007) specifically recruited separate clinical and control comparison groups, their studies each included a large community sample, adding strength to the study. For instance, Dunn and Sacco recruited 439 community dwelling older adults, with 36 meeting DSM-III criteria for depression. Similarly, Low and Hubley's study involved 119 cardiac patients, with almost 12% diagnosed with a depressive disorder, and Lewinsohn et al. utilized 1005 community based participants, with 8% of individuals determined to be depressed. By including a large sample size in their studies and confirming the presence of depression using a diagnostic interview, these researchers were able to create depressed and non-depressed groups that could be compared on the measures of interest. The large age range for both the Low and Hubley and Lewinsohn et al. studies also lend generalizability of the results for the population investigated. These findings are in-line with a systematic review of the criterion-based validation studies for the GDS, which found sample size to range from 40

to 715 individuals in 42 studies (Wancata, Alexandrowicz, Marquart, Weiss, & Friedrich, 2006).

Additionally, it is noted that, in conducting studies in an effort to provide criterion related evidence for a measure, several studies strengthened support for their results by controlling for specific variables that had the potential to confound results. Lewinsohn et al. (1997) controlled for cognitive and functional impairment and Low and Hubley (2007) and Snyder et al. (2000) screened for cognitive impairment using the MMSE. Several studies also screened for physical disease (Lewinsohn et al., 1997), central nervous system disease (Olin et al., 1992), or the use of psychotropic medication within two weeks prior to commencement of the study (Olin et al., 1992).

This review has highlighted that, in criterion validation, the use of a clinical and a community sample is occasionally utilized, although not necessary, as several studies used just a community-based sample from which different groups were created based on the results of a diagnostic tool. When a clinical sample is not specifically recruited for a study, the strength of the study can be increased with the use of a large sample size. Regardless of sample characteristics, the use of a recognized diagnostic tool to confirm the absence or presence of depression in the sample, and to serve as the criterion measure, appears to add considerable strength and credibility to study results. In the analysis of between-group differences, an appropriate t-test or F-test should be used if the data are normal and there is equality of variance between groups, although no studies commented on the normality or the equality of variance of the data between groups. If the data are non-normal, then an appropriate non-parametric test should be used. It also appears that an ROC curve analysis is an effective way to determine the sensitivity,

specificity, PPV, and NPV of a measure and provide further evidence of criterion validity for a given cut-off score.

2.5.4 Factorial validity evidence

Factorial analysis can be used to provide validation evidence for a measure by determining or confirming the underlying dimensions of a measure (Shultz & Whitney, 2005). That is, factorial analysis is used to examine if a scale is uni-dimensional or multi-dimensional and identify or reveal dimension(s) that are a facet of the construct that the scale is intended to measure. Shultz and Whitney (2005) described that factorial analysis reduces the number of factors from a maximum of the original number of items by assessing the intercorrelation among measure items. If items are correlated, it is the result of some dimension that they have in common. In cases where the underlying dimensions of the measure are not certainly known, exploratory factor analysis (EFA) can be used. Most EFA procedures divide the variance of items into common, specific, and error variance and extract factors based on the common variance of the original items (Shultz & Whitney, 2005). In cases where the dimensions are theoretically suspected and require confirmation, confirmatory factor analysis (CFA) may be used. In contrast, in CFA, researchers must specify the number of factors and specify which items will load on a particular factor (Shultz & Whitney, 2005).

Fewer factor analyses were revealed in the review as compared to studies providing criterion-related validation evidence. These studies include an investigation into the factorial validity of the GDS-5 (Marquez et al., 2006), the BDI-II (Segal et al., 2008; Osman et al., 2004), the CMDI (Nyenhuis et al., 1998), the Cognitive Checklist

(CCL; Beck, Brown, Steer, Eidelson, & Riskind, 1987), and the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), the latter of which were two measures used in a study by Beck, Novy, Diefenbach, and Stanley (2003) to differentiate anxiety and depression.

CFA was used to verify a single factor model that provided an excellent fit for the GDS-5 with a sample of 278 sedentary community-dwelling older adults (Marquez et al., 2006). The comparative fit index (CFI) in this study was found to be 1.0, indicating a very good model fit, since the value is greater than .90, which indicates a good model-data fit (Hu & Bentler, 1999). The authors reported that the model fit was also supported by a non-significant chi-square statistic and appropriate value for the root mean square error of approximation (RMSEA) of .05. In examining the appropriateness of the sample size for this study, it was found to be sufficient, with a sample size to number of items ratio being greater than 5:1 (Shultz & Whitney, 2005).

CFA was also used to confirm underlying factors of the BDI-II in a study utilizing a sample of 376 young and old community dwelling adults (Segal et al., 2008). Segal et al. showed that the two factors representing non-cognitive and cognitive dimensions that were found in previous studies with depressed geriatric inpatients (Steer, Ball, Ranieri, & Beck, 1999; Steer, Rissmiller, & Beck, 2000) did not fit the results from either the younger or older age groups within their community sample. The study results showed that the chi-square statistic was significant, with a low CFI of .69 and high RMSEA of .08. A subsequent principal component analysis (PCA), an approach to EFA that examines the total variance among variables and not a correlational matrix, revealed a satisfactory one-factor solution for both the younger and older age group. A single factor

for the younger group had an eigenvalue of 8.08, accounting for 38% of the variance, and a single factor for the older group had an eigenvalue of 6.35, explaining 30% of the variance. This study succinctly and clearly describes the process of gathering factorial validation evidence for a measure when administered to a different population, for which the factor structure is different than that originally proposed, with CFA performed first, followed by PCA. Again in this study, the sample size was found to be sufficient for the 21-item BDI-II.

Osman et al. (2004) investigated the factor structure of the BDI-II in a study involving 408 male and female psychiatric inpatients. The authors first utilized CFA to confirm two- and three-factor solutions for the BDI-II, as published in other studies with adolescent psychiatric inpatients, and also evaluated an alternative one-factor model. All results showed a significant chi-squared value and lack of good fit for the BDI-II. Osman et al. subsequently conducted EFAs to explore alternate solutions for the BDI-II items, using maximum-likelihood estimation with means and variance-adjusted chi-square (MLMV) to estimate the model. The two- and three-factor solutions had comparable fit estimates, but three items failed to load adequately on any of the three-factor solutions, so the two-factor solution was retained, with a cognitive-affective factor eigenvalue of 8.08 that accounted for 29% of the variance, and a somatic factor eigenvalue of 1.15 that accounted for 13% of the variance. In addition to the use of the eigenvalues >1 extraction rule, scree plots were also used to confirm the number of factors. This outcome is in contrast to Segal et al. (2008) who showed a one-factor solution for the BDI-II for a sample of community dwelling adults. Given that the Segal et al. study involved

clinically diagnosed adolescents, it is understandable that the factor solution may differ due to the difference in the sample population.

In an additional study using a sample of 83 older adults with generalized anxiety disorder, Beck et al. (2003) found an adequate two-factor solution for the CCL for depressive and anxious symptoms using CFA, but inspection of the residual correlations indicated a three-factor solution would improve model fit. An EFA using MLMV was subsequently conducted. An examination of the scree plot revealed that, in fact, a two-factor solution was appropriate and accounted for 49% of the variance in CCL scores. In the same study, CFA confirmed an adequate two-factor fit for the PANAS, although some of the factor loadings were small. Subsequently, the examination of a scree plot, generated by an EFA using MLMV, revealed a three-factor solution for the PANAS for positive affect, negative affect of anxiety and anger, and negative affect of guilt and shame, accounting for 55% of the variance in PANAS scores. This study indicates that EFA is an effective method for revealing the underlying factor structure for a measure when using it with a new population. In this case, it revealed that, with a sample of older adults with generalized anxiety disorder, a two-factor solution for the PANAS may not be as good a fit, and these results communicate issues with the construct validity of the measure with respect to a particular population. These results may have also been influenced by sample size. In examining the appropriateness of the sample size for this study, it appeared to be too small for the 26-item CCL and 20-item PANAS. Without an adequate sample size, results will be unlikely to generalize to other samples.

In contrast, in developing and initially validating the CMDI, Nyenhuis et al. (1998) used PCA to first perform an item analysis on the 42-item CMDI, based on results

from 454 community participants. This analysis revealed a five-factor solution as per scree plot criteria. CFA was subsequently performed on the scores from 420 alternate participants. Although a five-factor solution revealed a significant chi-square statistic, the authors noted that the high sample size could have produced statistical significance. Two other goodness-of-fit measures confirmed the five-factor solution, including the non-normed fit index (NNFI) and the CFI, with values of .92 and .93, respectively.

The review of studies that provided factorial validity evidence revealed that CFA is predominantly chosen for factorial analysis in the validation of an existing measure when used with a new population. Examination of these studies also revealed that, if the suspected number of factors of the measure were not confirmed with CFA, then follow-up EFA analysis was conducted to reveal the factor structure with the population of interest. Conversely, this would suggest that, when conducting a study to provide factorial validity evidence for a new measure, EFA would be conducted first, followed by CFA, as was the case in Nyenhuis et al. (1998). In the reviewed studies, when EFA was conducted, either PCA or MLMV was used. As well, in the studies that conducted an EFA, the number of factors extracted was determined by different methods: eigenvalues >1, scree plot, or both. It is also noted that samples used in the reviewed studies contained either clinical or community participants, but not combined samples, and the population sample sizes in four of the five studies were sufficiently high with respect to the number of items comprising the measure of interest.

2.5.5   Convergent and discriminant validity evidence

Convergent validity evidence is provided when a test correlates relatively highly with other variables or tests that measure the same or a related construct (Gregory, 2004). In contrast, discriminant validity evidence is provided when a test does not correlate relatively highly with variables or tests that measure unrelated or less related constructs (Gregory, 2004). In a review of depression measure validation studies, studies providing convergent and discriminant validity evidence were more numerous as compared to studies providing content, criterion, or factorial validity evidence. In the studies that were reviewed, a number of common depression measures were often either the focus of the study or were used as a convergent measure in the study.

Several studies only provided convergent validity evidence. Yesavage et al. (1983) examined convergent validity evidence for the GDS and Joseph, Lewis, and Olsen (1996) provided convergent validity evidence for the Depression-Happiness Scale (D-H S). Dunn and Sacco (1989) conducted a validation study of the GDS and SDS by correlating these measures with each other and with the DSC.

It appears that a majority of studies, however, examined both convergent and discriminant validity evidence, rather than solely investigating convergent evidence. It is also noted that, in many studies (Marquez, 2006; Osman et al., 2004; Segal et al., 2008; Yesavage et al., 1983), the analysis that was used to determine the relationships between measures, an estimate of the level of convergent and discriminant validity, was a Pearson product-moment correlational analysis. In one study though, Osman et al. (2004) conducted partial correlations to control for co-morbid anxiety related symptoms in a

study with older adults because of the suspected overlap in symptomatology between depression and anxiety in this population.

As a result of this review, it was also noted that the populations used in these studies were varied. A few studies (Beck et al., 2003; Osman et al., 2004; Snyder et al., 2000), with sample sizes of 54, 83, and 319, utilized a sample from which clinically depressed individuals were identified through diagnosis after recruitment, with the majority of the sample being non-depressed. Yesavage et al. (1983), in contrast, used a pre-diagnosed clinically depressed group and a non-depressed community group in their study, with a total of 100 participants. Only one study used an entire sample of 77 diagnosed clinically depressed individuals (Cahill, Barkham, Stiles, & Twigg, 2006). A majority of studies included just a community sample of adults (Beck et al., 2003; Dunn & Saaco, 1989; Joseph et al., 1996; Marquez et al., 2006; Osman et al., 2004; Segal et al., 2008), with sample sizes ranging from 194 to 439 individuals. It appears that, in studies using a community sample, the sample sizes were relatively large. Conversely, when a clinical population was used, sample sizes were lower, most likely due to the difficulty in accessing and/or diagnosing a clinical population.

Of additional importance in potentially guiding the procedures used in this study is to identify the measures used in providing convergent validity evidence for depression measures. As revealed by the literature review, these convergent measures are listed in Table 2.2, with each accompanied by the original measure citation, the construct measured, and the list of studies in which the measure was identified as providing convergent validity evidence. It is noted that the convergent measures are predominantly other depression measures, but some measures of related constructs such as pessimism,

suicide, and anxiety have also been used to provide convergent validity evidence. Scales of wellbeing also provided convergent validity evidence, with correlations that were inversely related to scores on the depression measure of interest.

Table 2.1 Measures Providing Convergent Validity Evidence

Measure Name	Measure Acronym	Measure Citation	Construct Measured	Study Citation
Beck Depression Inventory	BDI	Beck et al., 1961	Depression	Beck et al., 2003; Joseph et al., 1996; Snyder et al., 2000
Beck Depression Inventory-II	BDI-II	Beck et al., 1996	Depression	Cahill et al., 2006
Geriatric Depression Scale	GDS	Yesavage et al., 1983	Depression	Beck et al., 2003; Dunn & Sacco, 1989; Marquez et al., 2006; Snyder et al., 2000
Zung Self-Rating Depression Scale	SDS	Zung, 1965	Depression	Joseph et al., 1996; Dunn & Sacco, 1989; Yesavage et al., 1983
Center for Epidemiological Studies Depression Scale	CES-D	Radloff, 1977	Depression	Joseph et al., 1996; Segal et al., 2008
Hamilton Rating Scale for Depression	HRSD	Hamilton, 1960	Depression	Cahill et al., 2006; Yesavage et al., 1983

Measure Name	Measure Acronym	Measure Citation	Construct Measured	Study Citation
Coolidge Axis-II Inventory	CATI	Coolidge & Merwin, 1992	Depression and anxiety	Segal et al., 2008
Depression Symptom Checklist	DSC	Sacco, 1983	Depression	Dunn & Sacco, 1989
Beck Hopelessness Scale	BHS	Beck, Weissman, Lester, & Trexler, 1974	Pessimism	Osman et al., 2004
Suicidal Behaviors Questionnaire-Revised	SBQ-R	Osman et al., 2001	Suicide	Osman et al., 2004
State-Trait Anxiety Inventory	STAI	Spielberger, 1983	Anxiety	Beck et al., 2003
Beck Anxiety Inventory	BAI	Beck & Steer, 1993	Anxiety	Osman et al., 2004
Clinical Outcomes in Routine Evaluation-Outcome Measure	CORE-OM	Strupp, Horowitz, & Lambert, 1997	Wellbeing, functioning, and harm to self or others	Cahill et al. 2006
Short Psychological Well-Being Scale	SPWS	Ryff, 1989	Wellbeing	Segal et al., 2008
Satisfaction with Life Scale	SWLC	Diener, Emmons, Larson, & Griffen, 1985	Wellbeing	Marquez et al., 2006

Similarly, an overview of the identified discriminant measures used in the reviewed depression validation studies will also assist in determining the procedures used in the study. As identified by the literature review, measures that provided

discriminant validity evidence for depression measures are outlined in Table 2.3, with

each accompanied by the original measure citation, the construct measured, and the list

of studies in which they were identified as providing discriminant validity evidence.

Table 2.2 Measures Providing Discriminant Validity Evidence

Measure Name	Measure Acronym	Measure Citation	Construct Measured	Study Citation
Minnesota Multiphasic Personality Inventory-Adolescents (externalizing scales)	MMPI-A	Butcher et al., 1992	Anger and conduct problems	Osman et al., 2004
Inventory of Interpersonal Problems-Avoidant	IIP-A	Horowitz et al., 1988	Social desirability and lack of sociability	Cahill et al., 2006
Barriers Self-efficacy Scale	N/A	McAuley, 1992	Perceived physical ability	Marquez et al., 2006
Exercise Self-efficacy Scale	N/A	McAuley, 1992	Perceived physical ability	Marquez et al., 2006
Self-ratings of overall perceived health	N/A	N/A	Perceived health level	Segal et al., 2008
Beck Hopelessness Scale	BHS	Beck et al., 1974	Hopelessness	Cahill et al., 2006
State-Trait Anxiety Inventory	STAI	Spielberger, 1983	Anxiety	Snyder et al., 2000
Penn State Worry Questionnaire	PSWQ	Meyer et al., 1990	Worry	Beck et al., 2003; Snyder et al., 2000

Measure Name	Measure Acronym	Measure Citation	Construct Measured	Study Citation
Worry Scale	WS	Wisocki et al., 1986	Worry	Snyder et al., 2000
Quality of Life Inventory	QOLI	Frisch, 1994	Life satisfaction	Snyder et al., 2000

This review first revealed an overlap in the Beck Hopelessness Scale (BHS) as both a convergent and discriminant measure. Second, two anxiety measures, the STAI and the BAI, provided conflicting roles in depression measure validation studies, even though these scales are designed to measure the same construct. Specifically, the STAI acted as a convergent measure in one study (Beck et al., 2003) and a discriminant measure in another (Snyder et al., 2000), and the BAI provided convergent evidence in a third study (Osman et al., 2004). These two observations reveal the blurred line between the determination of a measure as a convergent or discriminant measure. This determination appears to be a matter of degrees, rather than a clearly prescribed all-or-nothing decision, and is relative to the correlations found between the measure of interest and other convergent or discriminant measures utilized in the study. For example, Cahill et al. (2006) indicated that partial support for the discriminant validity of the BDI-II was demonstrated through the relatively lower correlations of .34 and .45 with the more narrowly depression-related constructs of hopelessness and pessimism (as measured by the BHS and the IIP-Av, respectively). This is in contrast to the correlations of .67 and above with the convergent measures.

It is also noted that the measures shown to provide discriminant validity evidence for depression measures are based on a variety of disparate constructs. It

appears that many more measures have the capability of providing discriminant validity evidence for depression measures, as compared to those available to act as convergent measures.

This overview of convergent and discriminant measures also provides an indication of the actual numbers of measures that have been utilized in the reviewed studies to provide this type of validation evidence. It is noted that most studies utilized at least two convergent measures to provide evidence of convergent validity, with several studies incorporating three measures that provided this evidence. In contrast, the number of measures that provided evidence of discriminant validity was relatively low for most studies, if they were used at all, with most studies showing two measures that provided this evidence. The exception is the Snyder et al. (2000) study, in which four measures provided discriminant validity evidence.

Another outcome of the review that is relevant to this study relates to the significant and high correlations found between particular depression measures. Yesavage et al. (1983) demonstrated convergent validity for the GDS through large significant correlations between the GDS and the SDS and HRSD, with correlations of .80 and above between the measures. Snyder et al. (2000) demonstrated similarly high convergent validity between the GDS and the BDI. Osman et al. (2004) demonstrated that the BDI-II was significantly correlated to the BAI, SBQ-R, and BHS, although these correlations fell in the range of .51 to .69. Segal et al. (2008) demonstrated convergent validity of the BDI-II with high correlations between it and the CES-D, CATI (Depression and Depressive Personality Disorder subscales), and the SPWS, with moderate correlations of .58 to .68. In addition, Cahill et al. (2006) also revealed

correlations between the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) and the BDI-II and HRSD of .75 and .67, respectively, thus supporting convergent validity. This body of evidence points towards suspected convergent relationships among commonly seen depression measures including the GDS, BDI-II, SDS, HRSD, and CES-D, the knowledge of which may assist in the selection of a convergent measure, if this type of validation evidence is investigated in the study.

Further significant literature review findings are drawn from convergent and discriminant validation studies that included a co-morbid population to investigate the validity of depression measures with individuals also suffering from anxiety, a theoretically separate construct with overlap in symptomatology with depression (Cassidy, Lauderdale, & Sheikh, 2005). In particular, Snyder et al. (2000) conducted a convergent and discriminant validation study of the BDI and GDS with individuals diagnosed with generalized anxiety disorder (GAD) and affective disorders. Convergent validity was demonstrated by a significant positive correlation between scores for the BDI and the GDS. Discriminant validity was demonstrated by lower correlations for the BDI and GDS with the STAI, PSWQ, WS, and QOLI. Results supported the validity of the BDI and the GDS with a sample of anxious older adults. In a related study, Beck et al. (2003) evaluated convergent and discriminant validity of the CCL and PANAS in a study that focused on differentiating anxiety and depression in a clinical sample of 83 older adults meeting DSM-IV criteria for GAD. Findings demonstrated a high correlation between the anxiety and depression factors, indicating that, in older adults, the difference between the two constructs is clouded. These findings are supported by results from Segal et al. (2008) that showed a high correlation

between the BDI-II and the CATI-Anxiety subscale, indicating poor discriminant validity between depressive symptoms and anxiety-based symptoms.

From a study design perspective, several strengths and limitations were identified in the reviewed studies. For example, a number of studies (Marquez et al., 2006; Segal et al., 2008) provided large amounts of convergent validity evidence through correlations to several scales of similar constructs. In addition, Cahill et al. (2006) strengthened results of their convergent validation study of the CORE by incorporating study design elements that controlled specific variables. For example, all 77 participants were clinically diagnosed with a major depressive episode using a structured interview based on DSM-IV depressive symptomatology. As well, prior to the interview, the potential participants were also screened using the BDI-II to check for the presence of depressive symptoms and asked about the use of psychoactive medication. Although the authors noted that 62% of participants were on prescribed medication, they did not indicate if the use of psychoactive medication was one of their exclusionary criteria. A limitation in this study included the use of only a clinical sample; inclusion of a non-depressed community sample would likely have increased the variance of scores on the measures.

The validation evidence provided for depression measures gained from a convergent and discriminant validity perspective appears to be larger than the body of evidence gathered using other approaches to validity (i.e., convergent and discriminant validity evidence appears more often in studies as compared to content, criterion, or factorial validity evidence). It was also discovered that the samples used in studies included depressed individuals only, non-depressed individuals only, and both groups.

Moreover, sample sizes were often lower when a clinically depressed group comprised part, or the entirety, of the sample. This review has also highlighted that, in the analysis of evidence of convergent and discriminant validity, a blurred line might exist between the determination of a measure as a convergent or discriminant measure. Several study design strengths were also noted, including the use of numerous convergent measures that provide large amounts of convergent validity evidence through correlations to several scales of similar constructs (which were mainly other depression measures), and the variety in the constructs of the discriminant measures. Of note as well were specific study findings that demonstrated a high correlation between the anxiety and depression factors, indicating that, in older adults, the difference between the two constructs is blurred.

2.5.6   Reliability evidence

This review of depression measure validation studies concludes with a review of the reliability of the scores obtained on measures examined in these studies because, without reliability, a measure cannot be valid. The reliability of scores from a measure shows the ability to provide consistency in measurement (Gregory, 2004). More specifically, it is an estimate of how consistently a set of items measures a construct and the degree to which an instrument measures the same way each time it is used, under the same conditions, with the same individuals. Because validity requires reliability, reliability is also a form of validity evidence. Gregory (2004) explained, by way of example, that decisions based on unrepeatable test results are un-informed and unethical.

In the reviewed studies, evidence of reliability was provided in a number of ways. These methods included: a) internal consistency as measured by item-total correlations, b) internal consistency as measured by the intercorrelation among items, c) an overall measure of internal consistency provided by Cronbach's coefficient alpha, d) internal consistency as measured by split-half reliability, and e) test-retest reliability. The above listed estimates of internal consistency each provide a basis for judging the extents to which the scale's items all measure the same underlying construct (Yesavage et al., 1983). Cronbach's alpha splits the measure in two in every way possible, computes correlation values for every split, corrects for the shortened length of the test using a Spearman-Brown correction, and generates a single value for Cronbach's alpha whereas split-half reliability is calculated by correlating pairs of scores from two halves of a test and typically applying a Spearman-Brown correction (Gregory, 2004). The test-retest estimate of reliability, in contrast, provides an indication of the stability of the scores over time (Yesavage et al., 1983). Test-retest reliability is determined through the administration of a measure at two different times, to the same sample, and calculating the correlation between the two sets of results. Irrespective of the reliability estimate used, the closer these values are to 1.0, the higher the reliability estimate of the instrument.

Of the studies reviewed, only one study did not provide any estimate of reliability based on participant results (Olin, et al., 1992). Every study that provided reliability evidence (Beck et al, 2003; Cahill et al., 2006; Dunn & Sacco, 1989; Joseph et al., 1996; Lewinsohn et al., 1997; Low & Hubley, 2007; Marquez at al., 2006; Nyenhuis et al., 1998; Osman et al., 2004; Segal et al., 2008; Snyder et al., 2000;

Yesavage et al. 1983) provided an estimate of reliability using Cronbach's alpha. The Cronbach's alphas for the measures of interest in each of these studies ranged from .59 to .94 for the total scale score, with most measures possessing alphas in the .82 to .94 range, indicating good to high internal consistency. It is also noted that the majority of these studies also provided Cronbach's alphas for subscales, if they existed, as well as total scale alphas for distinct groups within the study (e.g., males and females, depressed and non-depressed individuals). Alphas were also generally calculated for convergent or discriminant measures used in a particular study.

A few studies also provided a measure of the internal consistency of a scale based on a median value of item correlations with the total score (Dunn & Sacco, 1989; Snyder, Stanley, Novy, & Averill, 2000; Yesavage et al. 1983), with values ranging from .32 to .56. These values are satisfactory item correlations with the total score and provide an indication that all items do measure a common latent variable. Similarly, Osman et al. (2004) and Yesavage et al. (1983) provided measures of internal consistency using mean item intercorrelation values, which ranged from .36 to .40. These values are also in a range that is acceptable to indicate a high degree of internal consistency. Item-total and inter-item correlation values are generally lower than Cronbach's alpha values because of the nature of correlations with the total score or with other items. It appears that this is not necessarily the optimum method by which to estimate internal reliability, but is best used to identify items that do not correlate with other items or with the total score for possible elimination from the scale.

It was also noted that a couple of studies provided split-half reliability values for the measure of interest (Nyenhuis et al., 1998; Yesavage et al., 1983), with values of

.84 and .94. As Cronbach's alpha can be viewed as the mean of all split-half coefficients corrected using Spearman-Brown correction (Gregory, 2004), it is expected that these values would roughly fall in the same range as those previously observed for Cronbach's alpha. As well, Lewinsohn et al. (1997) and Yesavage et al. (1983) provided test-retest reliability estimates for their measure of interest, with values of .52 and .85, respectively. Yesavage et al. used a test-retest interval of one week and Lewinsohn et al. did not indicate their test-retest interval.

Through this review of reliability-based validation evidence, it was observed that an overwhelming majority of studies provided reliability evidence based on the results of the measure derived from the sample used in the study. All of the studies that provided reliability evidence reported Cronbach's alphas, with most values falling in the range of .82 to .94, indicating sufficient internal consistency for the measure with that particular sample. It was also noted that alphas were often provided for subscales, sample groups, and other measures used in the study. It was revealed that most of the studies provided reliability evidence using one or two forms (e.g., Cronbach's alpha and inter-item correlations with the total score), with only one study (Yesavage et al., 1983) providing up to five forms of reliability evidence.

2.5.7   Summary

The process of validation is always ongoing. Therefore, no single study can establish validity. Through this lens, it was observed that a number of studies included multiple validation approaches to build a body of evidence to support the inferences of scores of a measure used with a particular population. It also appears that a large

proportion of studies included criterion, as well as convergent and discriminant validation evidence as part, or comprising the entirety, of the study. Moreover, the literature review revealed that the GDS, BDI-II, CES-D, and SDS are commonly used depression measures and they have often been examined, either as the focus of a validation study or as a convergent measure for another depression measure. Noted strengths of several of the studies include the use of a diagnostic tool to confirm the presence or absence of depression in a sample, as well as the use of both a clinical and community based sample. Moreover, assessment of cognitive functioning to identify cognitive impairment and, in at least one study, the exclusion of participants using psychoactive medication assisted in controlling for these variables.

Recommendations stemming from the preceding literature review therefore include the provision of convergent and discriminant validity evidence for the HDS-OA using multiple measures, and the use of both clinically depressed and non-depressed community samples that have a confirmed diagnosis of depression or a confirmed absence of depression gained through a current diagnosis or use of a diagnostic tool. With the use of two groups such as these, criterion-related validity evidence can also be provided. It is also recommended that the convergent measure(s) be another measure of depression and the discriminant measure(s) can be a scale that measures a less related or non-related construct, given that discriminant validity is provided by degrees of association as compared to a convergent measure. It is also recommended that the reliability of HDS-OA scores and any convergent or discriminant validity measure scores be examined using Cronbach's alpha as an indicator of internal consistency. The

following section will elaborate on these recommendations and focus on the rationale for the selection of validation procedures used in this study.

2.6   HDS-OA Validation

No validation studies of the HDS-OA have been published to date, although data have been collected in two studies to examine reliability and criterion validity evidence of the HDS-OA. Convergent validity evidence using the BDI-II as a convergent measure was also examined in both of these studies. When considering the selection of validation procedures and the design of this study, the approaches used in previous depression measure validation studies guided the decisions made. As revealed in the literature review, many validity studies of other depression measures involved several types of validation procedures. It was noted particularly that the majority of these studies focused on criterion and/or convergent and discriminant validity, with a large proportion of the research involving the latter, conducted independently or in conjunction with another type of validation effort. It was also evident that many validation study designs included both depressed and non-depressed participants, with a confirmed clinically depressed group and a confirmed non-depressed group strengthening the results of the study.

Considering that the intent of the present study was to provide validation evidence that would support inferences made from the HDS-OA scores and to investigate whether the HDS-OA is as "useful" a tool for the screening of depression in older adults as compared to another depression screen validated for use with older adults, a combination of validation approaches was selected. This decision was consistent with the objectives of the study as well as with the unitary approach to test validation. Given the findings of the

literature review, it was determined that the study would investigate the reliability of the scores of the HDS-OA, evaluate criterion validity, and provide an estimation of the convergent and discriminant validity of the measure. To achieve these objectives, I incorporated a group of already clinically diagnosed depressed older adults and a confirmed non-depressed group of community-based older adults in order to investigate criterion-related validity, selected several measures for inclusion in the study in order to provide estimates of convergent and discriminant validity, and chose the GDS as geriatric depression measure of comparison to the HDS-OA.

As reviewed, other types of validation work could possibly have been pursued. Because the HDS-OA is a relatively new measure though, with limited validation evidence collected thus far, a reliability, criterion-based, and convergent and discriminant construct validation study was appropriate at this time. These approaches have been supported in the research literature related to the validation of depression measures and conducted with a variety of populations. Due to time and monetary constraints, for example, a factorial validation approach could not easily be undertaken due to the number of participants required by this design, which would need at least 80 to 160 participants, based on the number of items of the HDS-OA. This would be onerous to complete in terms of recruitment of a clinically depressed sample or in terms of confirming the absence of a current major depressive disorder in a community-based sample using a diagnostic interview. No one study can provide all necessary evidence to support the validity of a measure, so selectivity in approaches was required.

In order to provide an estimate of the reliability of the HDS-OA, Cronbach's alpha was selected as a measure of internal consistency, as this approach was used most

often in the reviewed studies. It was decided that the study would not investigate test-retest reliability because it would double the time requirements to administer measures and also impose an additional time burden on study participants.

This study was also designed to investigate the ability of the HDS-OA to accurately detect the presence or absence of the depression in the sample being screened using a criterion validation approach. The use of two groups in the study design, composed of individuals with a diagnostic confirmation of the presence or absence of depression, facilitated the ability to investigate this type of criterion validity. In a systematic review of GDS criterion-based validation studies, studies were only selected for inclusion in the review if a "gold standard" research interview was used to diagnosis cases (Wancata et al., 2006). To estimate the ability of the HDS-OA to correctly identify cases of depression as well as correctly identify those who are not depressed, ROC curve analyses was selected as the method to calculate sensitivity, specificity, PPV, and NPV of the measure based on different cut scores, with the AUC providing an overall estimate of the accuracy of the measure. In the present study, the optimum cut score selected was the one that weighed more heavily on indentifying true cases of depression (i.e. sensitivity), rather than accurately identifying truly non-depressed cases (i.e. sensitivity). These values were compared to those obtained with the GDS, which was also be administered to participants, in order to determine if the HDS-OA was at least as equally good a measure in screening for depression in older adults. In addition, a between groups comparison was selected for inclusion in the statistical analysis to provide further evidence of criterion validity by comparing mean scores between the depressed and non-depressed groups in order to detect if there was a statistically significant difference

between the groups (i.e. known-groups validity). This would provide an indication of the ability of the HDS-OA to differentiate between depressed and non-depressed individuals. Results would also be compared to those for the GDS.

In addition, the study was designed to provide evidence of convergent validity based on the correlation of scores between the HDS-OA and those for the GDS as a convergent depression measure to demonstrate that the HDS-OA was measuring the construct of depression. It was decided that only one separate measure would be used as a convergent measure in this study, despite the evidence provided by the literature review that showed several different convergent measures are often used to provide evidence of convergent validity. This decision was made to reduce the time required for measure administration (i.e., reduce subject burden) and was also made based on the fact that there is currently only one commonly used self-report depression measure that has been designed to screen for depression in older adults. In addition, the GDS is a measure that uses a yes-no response format, similar to the HDS-OA. For these reasons, the GDS was selected as the convergent measure of focus in the present study because it provided a similar measure of the same construct to which to compare the HDS-OA. In further support of its use, the body of validation evidence for the original 30-item GDS is large and the measure has been used to screen for depression in numerous general older adult populations and diverse older medical populations (Rapp, Parisi, Walsh, & Wallace, 1988). The fact that the GDS has been used with older medical populations is particularly relevant in the present study because the depressed group in this study was recruited from hospital- and clinic-based elder care programs in order to include subjects with a confirmed a diagnosis of depression. The commonly used BDI-

II was not selected as a convergent measure because it was not designed to screen for depression specifically in older adults and it has already been used as a convergent measure in two other HDS-OA validation studies.

Although the majority of the depression measure validation literature focuses on the original 30-item GDS, it was discovered after the study was designed and at the outset of the research that the hospitals and health centers, from which depressed subjects for this study were recruited, all used the short-form GDS-15 (Sheikh & Yesavage, 1986) to screen for depression in older adult patients. It appears that this usage is not specific to Vancouver because comments were made by clinicians and researchers at the 2009 American Psychological Association convention in Toronto, Ontario with reference to the use of the GDS-15 in clinical settings and its applicability as a convergent measure to which the HDS-OA could be compared (A. Hubley, personal communication, September 10, 2009). Items 1, 2, 3, 4, 7, 8, 9, 10, 12, 14, 15, 17, 21, 22, and 23 of the original 30-item GDS comprise the GDS-15 (Sheikh & Yesavage, 1986). Given that the 30-item GDS and the GDS-15 showed similar sensitivity and specificity in a review of GDS criterion validation studies (Wancata et al., 2006) and the GDS-15 appears to be widely used in clinical practice, it was also of interest in the present study to investigate how these 15 items of the GDS performed as compared to the equivalent length 16-item HDS-OA. Therefore, scores on these items were extracted from the results and analyzed in the same manner and compared to the HDS-OA in the same way as the data from and GDS. It is noted that from this point forward, these 15 items of the GDS are referenced as the GDS-15 and the 30-item GDS will continue to be referred to as the GDS. Although caution must be taken in the

present study when making inferences from the results of the GDS-15 because this measure was not administered as a separate questionnaire, these collateral findings may suggest the extent of the similarities of the validation evidence for the HDS-OA and GDS-15 and assist in identifying areas of future research. Neither the shorter GDS-10 nor the GDS-5 were not selected for use in this study due to the few numbers of validation studies found in the research literature for these measures.

Finally, the study was also designed to determine the ability of the measure to discriminate between the construct of depression and different constructs in a sample of older adults by correlating scores between the HDS-OA and measures of less related constructs, thus providing evidence of discriminant validity. The MMSE was selected for use as a discriminant measure because it could be administered to study participants to identify cognitive impairment and the resulting cognitive status scores provided a measure of a different construct. As well, the study incorporated another measure of discriminant validity for the HDS-OA that was provided through a correlation with self-reported health scores collected with other demographic information. The BAI, an anxiety-screening tool, was also chosen to help provide a measure of discriminant validity. An anxiety screen was selected because the literature review revealed an overlap in symptomatology between anxiety and depression in older adults and, therefore, an investigation of the ability of the HDS-OA to differentiate between depression and anxiety was warranted and would be revealed by relatively lower correlations with the BAI as compared to the convergent measure (i.e., the GDS). This correlation would also be compared to that of the correlations between the GDS and the BAI to assess if the HDS-OA shows greater discriminant validity with the BAI. Despite the issues that older

adults have with the multiple-choice format of the measures like the BAI (Olin et al.,

1992), it was noted that this measure was carefully constructed to avoid confounding with

depression (Beck, Epstein, Brown, & Steer, 1988). This is a particularly important aspect

when using an anxiety inventory as a discriminant measure for a depression inventory

validation study due to the noted overlap in symptomatology between depression and

anxiety. Further reasons for the selection of the BAI are outlined below.

In deciding to use the BAI versus another commonly used anxiety measure, the

STAI, a study conducted with a population of older adults was of particular relevance.

The BAI and the STAI were both administered to 217 older adult outpatients with mixed

psychiatric disorders in order to determine the psychometric properties and diagnostic

utility of the BAI with older adults (Kabacoff, Segal, Hersen, & Van Hasselt, 1997). The

authors reported that both measures demonstrated high internal consistency reliabilities.

Specifically, the BAI demonstrated good factorial validity showing somatic and

subjective anxiety factors. In contrast, the authors found that the STAI did not

demonstrate factorial validity, resulting in a lack of support for the presence of state and

trait anxiety factors, upon which the measure is based. Both the BAI and Trait Anxiety

scale of the STAI demonstrated discriminant validity in separating patients with a current

anxiety disorder from patients without such a disorder. However, the State Anxiety scale

of the STAI did not discriminate between these groups. The authors concluded that the

results suggest that both the subjective factor and total score on the BAI can be useful as

a quick screening instrument in detecting presence of a current anxiety disorder in older

adult psychiatric outpatients. In addition, a subsequent study involving 40 older adults

investigated the suitability of using self-report measures of anxiety with older adults with

current or previous anxiety symptoms (Dennis, Boddington, & Funnell, 2007). Each participant received an independent diagnostic assessment and a rating of anxiety severity, and then completed the BAI, STAI, Hospital Anxiety and Depression Scale (HAD-S), and Visual Analogue Scale (VAS). The authors noted that participants made a high number of errors when responding to the STAI-Trait scale items and performed poorly in both screening and measuring severity on the Visual Analogue Scale (VAS). These findings resulted in the authors concluding that the BAI and HAD-S are the most acceptable measures for older adults. As well, in a study of the measurement of anxiety in community dwelling older adults, Fuentes and Fox (2000) reported that the STAI-T correlated more highly with the BDI, a measure of depression, than with other measures of anxiety, raising questions regarding the scale's discriminant validity and use as a general measure of anxiety, especially in older women.

The preceding discussion outlined the rationale for the decisions made with respect to the validation procedures selected, the resulting overall design of the study, and the inclusion of measures. The validation approach selected was based on the objectives of this research and on prior research conducted in the area of depression measure validation. The implementation of these choices is discussed in detail in the following chapter, entitled "Methods".

2.7   Hypotheses

Given the validation procedures and measures selected for this study, the following outcomes were hypothesized. It was anticipated that the HDS-OA would (a) produce reliable scores with an alpha greater than .80, (b) show stronger correlations with

the GDS and GDS-15 than the BAI and especially the MMSE and self-rated physical health item, (c) show a statistically significant difference between the mean scores of the depressed and non-depressed groups, (d) show adequate levels of sensitivity, specificity, PPV, and NPV, and (e) would produce equivalent or better performance as compared to the GDS and GDS-15.

Chapter 3 - Methods

3.1   Participants

Participants consisted of a total of 36 individuals, aged 60 and above. Eighteen

individuals were diagnosed with a MDE, dysthymia, or depressive disorder not otherwise

specified (NOS) at the time of measure administration. Individuals with depression due to

a general medical condition (GMC) or substance use (SU) or those experiencing

psychoses were excluded from the study. Another 18 individuals were non-depressed

community-based participants who were of similar ages and exact genders as compared

to the depressed group. All subjects were required to have a cognitive status score of 18

or greater on the MMSE.

Both a depressed group and a non-depressed group were included in the design of

the study in order to sample a range of depression severity, rather than restrict the sample

to solely a clinical or community sample, since the HDS-OA is intended to be used across

depression severity levels to screen for depression. The use of the two groups also

enabled a measure of the ability of the HDS-OA to distinguish between depressed and

non-depressed subjects.

In addition, the study was encompassing of the diversity of older adults in terms

of gender, age over 60 years, marital status, educational level, and socioeconomic

status, and did not control for these variables. The study was designed to be inclusive of

all ethnicities as well, although a limiting factor related to this were the additional

participation criteria of being able to speak, read, and write fluently in English.

3.2   Recruitment

In the study design, I required that the clinical population already have a current diagnosis of MDE, dysthymia, or depressive disorder NOS; therefore, it was decided that recruitment would best occur through clinical sources. Depressed group participants were recruited with the assistance of physicians, nurses, psychologists, and social workers from the (a) Geriatric Psychiatry Outpatient Program and Short-Term Treatment and Assessment unit at Vancouver General Hospital (Vancouver), (b) Geriatric Outpatient Clinic at Mount Saint Joseph Hospital (Vancouver), (c) Elder Care Program at St. Paul's Hospital (Vancouver), (d) North Shore Older Adult Mental Health Team associated with the West Community Health Centre (West Vancouver), and (e) Transitional Care Unit at Lion's Gate Hospital (North Vancouver).

Staff within participating geriatric health care centers identified potential depressed participants according to the study inclusion and exclusion criteria. To maintain confidentiality of patient information, these potential participants were initially approached by health care staff about the study and were provided with an information sheet and consent to be contacted form that described the research and participation requirements (see Appendix A). If individuals were interested in the study, they could contact me or consent to allow me to contact them by leaving their name and phone number on the form, and I then contacted them within one week to further discuss participation in the study. Upon contact, if the individual agreed to participate, we decided upon a time to meet in their own home or at the University of British Columbia (UBC) at a time that was convenient for them in order to administer the measures. I was also available to conduct in-person recruitment at geriatric outpatient clinics. For

example, interested individuals who were referred by a physician to the study during a clinic visit completed a consent form and were administered the measures during their hospital visit.

The community-based sample was recruited at the UBC Ageless Pursuits Lecture Series in June 2009 and in-person outreach to community-based senior centers and community centers, supplemented by advertisements (see Appendix B) posted in these locations.

3.3   Measures

The measures used in this study, the HDS-OA, GDS, BAI, and MMSE, are copyrighted measures and are not included in the appendices. It is noted though that the GDS is a freely available measure. The HDS-OA is also a freely available measure that is accessible through Dr. Anita M. Hubley, professor in the Department of Educational and Counselling Psychology, and Special Education at UBC. Similar to the measures, the SCID-I/NP interview is not included in the appendices, as it is a copyrighted and purchased diagnostic interview.

3.3.1 SCID-I/NP

The SCID-I/NP (First et al., 2007) is a diagnostic measure used to identify mood disorders in a non-patient population. It is a semi-structured interview based on the DSM-IV-TR and was constructed to assist researchers and clinicians in making diagnoses of Axis-I mood disorders in adults (First et al., n.d.). First, Gibbon, Spitzer, and Williams (2002) noted that the SCID-I/NP is designed for use in studies that do not

involve psychiatric patients, and thus was only used as the diagnostic interview for the community-based sample. The manual notes that the interviews generally can take from 60 to 90 minutes. The length of the interview is dependent on interviewee responses and takes from 40 to 75 minutes with non-depressed individuals; this estimate is based on 30 hours of training on the SCID-I/NP administration and scoring and interview experience that I gained during Summer 2008, and through experience gained during this research study.

A number of studies have used the SCID as the "gold standard" in determining the accuracy of clinical diagnoses (e.g., Shear, Greeno, Kang, & Ludewig, 2000; Steiner, Tebes, & Ludewig, 1995). In the present study, the SCID-I/NP was selected for use to confirm the absence of a current mood disorder, which includes major depressive disorder, bipolar disorder, dysthymic disorder, depressive disorder not otherwise specified, or mood disorders due to a general medical condition or substance use, and to confirm the absence of psychotic symptoms in individuals in the community based non-depressed sample.

3.3.2 MMSE

The MMSE (Folstein, & McHugh, 1975) is a short, clinician administered, easily scored, level A, 11-item questionnaire that focuses on the cognitive aspects of mental functioning. Total scale scores range from 0 to 30. In a review of the MMSE, Albanese (2003) indicated that the measure should be administered in a private environment and be provided by individuals who are trained in the administration of behavioral measures and clinical interviewing. The author noted that individuals with

relevant training in cognitive mental status assessment should interpret the results. I received training in the administration and scoring of the MMSE in May/June, 2008.

Albanese (2003) indicated that a score of 24 is an accepted cut score, with scores of 23 or lower indicating potential cognitive impairment that warrants further evaluation. The reviewer also reported that validation studies of the MMSE found sensitivities of at least 87% for a score of 23 or lower, a false negative rate of a maximum 13%, and a positive predictive value of at least 79%, providing good validity evidence. The MMSE Clinical Guide reports internal consistency alpha values ranging from .31 to .96, with values being higher among clinical samples due to increased variability in scores (Folstein, Folstein, & Fanjiang, 2001). The guide also reports test-retest coefficients results of .79 to .98 for studies involving depressed individuals, with test-retest intervals ranging from one day to 14.6 months.

In this study, the MMSE was selected for use to identify participants with cognitive impairment, as it is a commonly used measure of cognitive functioning in older adults (Albanese, 2003). A cut score of 18 was selected to identify individuals with moderate to severe cognitive impairment. Any individuals who scored less than 18 were excluded from the research to avoid invalidating scores on the other measures due to moderate to severe cognitive impairment. The relatively low cut score of 18 accommodated for lower education levels, increasing levels of cognitive impairment in older adults, and the tendency for depressed individuals to show less sustained effort in making cognitive decisions (Folstein, Anthony, Parhad, Duffy, & Gruenberg, 1985). The MMSE was also used as a discriminant measure in this study.

3.3.3   GDS

The GDS (Yesavage et al., 1983) is a level A, 30-item, self-administered measure of depressive symptoms that was specifically designed to screen for depression in older adults. A yes-no response format is used, and the authors indicated that the focus of the questions is based upon affective and cognitive rather than somatic symptoms that have been experienced over the past week. The scale is not based on DSM diagnostic criteria (Yesavage et al., 1983). Total scale scores range from 0 to 30, with higher scores denoting increasing severity of symptomatology and a conventional cut score of 10 (Rapp et al., 1988) or 11 (Brink, Yesavage, Lum, Heersema, Adey, & Rose, 1982). The GDS contains 10 reverse-scored items (items numbers 1, 5, 7, 9, 15, 19, 21, 27, 29, and 30). The measure is easily scored using an answer-key, with responses in concurrence with depressive symptomatology each receiving a score of 1.

Dunn and Sacco (1989) reported an alpha value for the GDS of .91, indicating high internal consistency. Similarly, in a paper on the development and initial validation of the GDS, Yesavage et al. (1983) reported an excellent internal consistency of .94 for the scale, and total scores on the GDS were reliable over a one-week period. Yesavage et al. also reported that evidence for the scale's validity was derived by comparing the mean scores from subjects who were identified as normal, mildly depressed, or severely depressed based on Research Diagnostic Criteria (RDC) for depression. The means of the three groups were shown to be reliably different.

3.3.4   HDS-OA

The HDS-OA (Hubley, 1998) is a 16-item, self-administered measure, designed for use as a screen for depression in adults aged 55 and older. The HDS-OA is also comprised of two additional (non-scored) questions used to capture information about current bereavement and new medication use. Questions are based on DSM-IV-TR criteria for major depressive disorder and therefore contain some somatic items. A yes-no response format is used for all questions, and each question contains a reminder of the previous two-week time frame within which to frame the answer as it has been found that respondents tend to forget the time-frame for questionnaire items within the first five or so items (A. Hubley, personal communication, July 14, 2008). The scale scores can range from 0 to 16, with higher scores denoting increasing severity of symptomatology. A cut score of 3 and higher was recommended based on preliminary results of a recent validation study of the HDS-OA for middle-aged and older adults (Hubley et al., 2009). The HDS-OA contains four reverse-scored items (item numbers 1, 3, 7, and 12). The measure is easily scored using an answer-key, with responses in concurrence with depressive symptomatology each receiving a score of 1.

3.3.5   BAI

The BAI (Beck & Steer, 1993) is a 21-item, multiple-choice, self-report, level B inventory used to measure the severity of an individual's anxiety. It consists of questions asking how the subject has been feeling during the past week, which are each rated on a four-point scale ranging from 0 to 3, for a total score that can range from 0 to 63 (Beck, Epstein, Brown & Steer, 1988). Beck and Steer noted that it is a quickly self-

administered measure that is easy to score and interpret. The BAI possesses the following cut points for anxiety: 0 to 7 = minimal, 8 to15 = mild, 10 to 18 = mild to moderate, 16 to 25 = moderate, 26 to 63 = severe (Beck & Steer, 1993). According to Beck et al. (1988), the BAI measures both the physiological and cognitive symptoms of anxiety, and item overlap with self-report depression inventories is minimized because it was designed to discriminate anxiety from depression.

Reliability and validity data were initially based on only three studies: the first used a mixed diagnostic group, the second used patients diagnosed with DSM-III-R anxiety disorders, and the third used a non-clinical sample (Dowd, 1998). Beck and Steer (1993) noted that the BAI was developed with psychiatric subjects and interpretation of results should be done cautiously with non-clinical individuals. Given this caution, the BAI is a measure that is appropriate for the age range of the population of interest, as it has been validated with older adult psychiatric outpatients (Kabacoff, Segal, Hersen, & Van Hasselt, 1997), and has more recently been validated for use with older adults with anxiety symptoms (Dennis et al., 2007).

The manual reports very good internal consistency reliability alpha coefficients, ranging between .85 and .94. Test-retest reliability data from Beck et al. (1988) showed a coefficient of .75 over a one-week interval. The validity evidence is derived from content, concurrent, construct, discriminant, and factorial validation studies, and generally shows excellent validity (Dowd, 1998).

3.3.5   Demographics form

In order to describe the sample, all participants were asked to complete a demographics form to provide information on age, date of birth, ethnicity, marital status, education level, self-rated health status as per the self-rated current overall health status item of the MOS 36-item short form health survey (SF-36; Ware & Sherbourne, 1992), and current anti-depressant medication usage (see Appendix C).

3.4   Controls in the Research Design

In order to test the hypotheses and establish criterion validity, two groups that differed in the presence of depression were included in this research. As such, there was an inability to randomly select and randomly assign participants to groups, given that the independent variable was pre-existing. The research design was, therefore, a quasi-experimental non-equivalent control group design study involving a clinically depressed group and a confirmed non-depressed community control group. The independent variable was the depression status of the participants and the dependent variable was the participant's scores on the measures of interest.

To control as much as possible for the differences between the two groups, the depressed and non-depressed group participants were equated to a degree by matching based on gender and age variables. Gender was matched exactly, but age was matched within a 5-year range (i.e., 2.5 years younger or older than the depressed group participant being matched), rather than by exact age, to allow for ease of recruitment of the community-based non-depressed group participants. Although the study controlled

for gender and age, it did not control for exact age, or for education, socioeconomic, or ethnicity variables.

The study was also designed to limit several other aspects of threats to internal and external validity. For example, I conducted all SCID-I/NP interviews with the non-depressed group to limit experimenter effects and trained research assistants administered the other measures (MMSE, HDS-OA, GDS, BAI) to the non-depressed group so we were blind to each other's results. In a systematic review of GDS criterion-based validation studies, Wancata et al. (2006) only included studies in which the diagnostic interviewer was blind to the results of the screening measures. For the already diagnosed depressed group, I administered the measures. Diagnostic interviews and the measures administered to individuals followed standard administration procedures and were not modified, thus limiting instrumentation threat. Measures were administered only once, which limited multi-treatment interference. As well, the administration of the GDS, MMSE, HDS-OA, and BAI to both groups was counter-balanced using a digram-balanced approach, in which each measure immediately precedes and follows another measure once (Keppel & Wickens, 2004), in order to limit practice or fatigue effects due to order effects. Participants in each group were then randomly assigned to one of four orders of presentation for the self-report measures and the MMSE using the Stat Trek random number generator program. This program can be accessed at http://stattrek.com/Tables/Random.aspx.

3.5   Procedures

3.5.1   Phase I: Selection and data collection - depressed group

A large proportion of individuals in the clinically diagnosed depressed group were recruited first. Individuals aged 60 or greater who had a current diagnosis of depression (i.e., MDD, dysthymia, or depressive disorder NOS) and were willing to volunteer, were invited to participate in the study. Primary hospital or health clinic contacts provided eligible individuals with an information sheet that outlined the purpose and requirements of the study and a place to leave their telephone number if they were interested in participating (see Appendix A). Individuals who provided contact information were then contacted within a week to discuss participation. When an individual agreed to participate, a meeting was scheduled, and a study session was conducted in a private location in the hospital or clinic during a visit (if their schedule and energy level allowed), in the individual's home, or at a private office at UBC, depending on the participant's preference.

During the session, informed consent was first explained to the participant. The consent form outlined the purpose of the study and participant rights and responsibilities (see Appendix D). If the individual agreed to the terms of the study, he/she was asked to sign the informed consent form, which was kept by the researcher, and the participant was given a copy for his/her records. After informed consent was obtained, the measures (HDS-OA, GDS, BAI, and MMSE) were then administered according to the guidelines for each scale. Finally, in order to describe the sample, the participant was also asked to complete a demographics data sheet (see Appendix C). When the session was complete, a

gift certificate was provided to the individual as a token of appreciation for their participation in the research.

The consent form for the clinically depressed participant group included an additional clause indicating consent for release of medical information recorded by the treating physician or psychologist, which pertained specifically to confirming conditions of depression and the absence of psychosis (see Appendix D). Therefore, the treating physician's name was also recorded during the session so that the participant's diagnosis of depression and the absence of a psychotic disorder could be confirmed by the treating physician or by another health care provider for that individual who had access to the subject's medical records after the session was completed. In many cases, this post-check was only required to confirm if there was a specific diagnosis of depression for the individual recorded on the medical chart (i.e., MDE, dysthymia, or depressive disorder NOS) because the health care professionals who had referred individuals to the study were aware of the study inclusion and exclusion criteria and only referred patients who met study requirements. In a majority of cases, the diagnosing physician did not indicate the severity level of the individual's depression and only recorded a general diagnosis of "depression" on the patient's medical chart. In these instances, the subject's diagnosis for this study was also simply recorded as "depressed".

Once all data were collected, the measures for each participant were assigned the same ID number, scored, and returned to a locked room at UBC.

3.5.2 Phase II: Selection and data collection - non-depressed community group

Individuals in the non-depressed community group were recruited and invited to participate in the study based on the genders and ages of subjects in the depressed group. The ages of the community sample and the clinical sample were not matched exactly but fell within an age difference of up to 2.5 years older or younger than the age of the depressed participant to lessen recruitment constraints. Participants were recruited according to the methods outlined previously and were screened in-person or via telephone for gender and age requirements that matched them to the clinically depressed group subjects.

Administration procedures remained the same for this group as those described for the clinically depressed group. In addition, individuals were provided the Referral List of Community Resources information sheet (see Appendix E) subsequent to obtaining signed consent. The consent form differed slightly from the depressed group because the non-depressed group did not need a diagnosis to be confirmed by a physician, psychologist, or psychiatrist (see Appendix F). I administered the SCID-I/NP interview and gathered demographic information immediately after a trained research assistant had completed informed consent and administered the questionnaires to the participant. These participants also received a gift certificate as a token of appreciation.

3.5.3 Materials required for the test session

The following materials were required for each test session, in order of administration: (a) Consent Form, (b) Referral List of Community Resources (for non-depressed community group participants), (c) study measures (HDS-OA, GDS, BAI,

MMSE administered in counter-balanced order) (d) SCID-I/NP Scoring Sheets (for non-depressed community group participants), (e) SCID-1/NP Summary Score Sheets (for non-depressed community group participants), and (f) Demographics Information Sheet.

Additional materials used in the sessions included: (a) pen (to sign consent form), (b) one extra copy of the consent form (for the subject to keep), (c) pencils and erasers, (d) one extra blank piece of paper for the sentence writing and drawing questions of the MMSE, (e) SCID-I/NP administration binder, and (f) gift certificate for participant.

3.5.4   Bookings

If a session with a participant was scheduled to take place in their home, their name, address, and phone number was collected during a phone conversation, and the participant was given my contact number (the Adult Development and Psychometrics Lab: 604-822-5250). If the session was occurring at UBC, a phone number and name was collected and the participant was provided information on how to get to UBC and directions to an easily identifiable meeting place on campus where I met the subject. The Adult Development and Psychometrics Lab served as the primary location for on-campus sessions, but other rooms were booked and used for the sessions if the lab was not available. When an appointment was booked, the day, date, and time of the appointment was confirmed at the end of the conversation, and the participant was provided a reminder call the day before the scheduled session.

3.5.5   Safety protocol

To ensure safety of the research assistants and myself during sessions that took place in a participant's home, these individuals accompanied each other to the session. In the cases where the research assistant left after the questionnaires were administered, I called the research assistant when the session was over to confirm that I had left the participant's home. When travelling alone to meet depressed individuals in their home, I contacted a study research assistant before and after the session.

3.5.6   Post-session inclusionary/exclusionary criteria

To meet criteria to be in the depressed sample, participants must have scored 18 or higher on the MMSE. Non-depressed community sample participants must also have met this criterion, as well as not be diagnosed (according to the SCID-I/NP) with a current mood disorder (i.e., MDD, bipolar disorder, dysthymic disorder, depressive disorder NOS, or a mood disorder due to GMC or SU) or have experienced psychotic symptoms. Individuals diagnosed with a current mood disorder (other than MDE, dysthymia, or depressive disorder NOS) and those experiencing psychotic symptoms were to be excluded from the study. After a session, if it was revealed that a potential community group match was diagnosed with an MDE, dysthymia, or depressive disorder NOS (as determined by the SCID-I/NP) that person's data were added to the depressed group and a non-depressed match for that individual was recruited. The presence of a past MDE did not automatically exclude individuals from the study, but the study avoided recruiting individuals with many recurrent and recent MDEs.

3.5.7   Research timeline

The ethics process began in February 2009. When ethics approval was obtained in mid-April 2009, recruitment of the clinically depressed group began. Recruitment and measure administration sessions occurred concurrently as research participants were identified. This phase of the research was of approximately five months in duration and resulted in the recruitment and testing of 18 depressed subjects.

Once a portion of the depressed group participants had been recruited and tested, the non-depressed group participants were recruited, beginning in mid-June 2009, as the required genders and age ranges of several non-depressed participants was known. As with the depressed group, recruitment and measure administration and interview sessions occurred concurrently as research participants were identified. The recruitment and testing phase for this group was of approximately three months in duration and resulted in the addition of 18 matched non-depressed subjects to the study.

Data were entered into SPSS in conjunction with recruitment, measure administration, and interviewing. The available data were then analyzed, and the results and discussion sections of this thesis were developed. These tasks took approximately two months to complete. This timeline is graphically depicted in Figure 3.1.

Figure 3.1 Research study timeline.

Research Task	Start Date	End Date	2009								
			J	F	M	A	M	J	J	A	S
Ethics application	02/01/09	04/13/09									
Depressed group recruitment	04/15/09	09/15/09									
Measure administration (depressed group)	05/01/09	09/15/09									
Non-depressed group recruitment	06/15/09	09/15/09									
Measure admistration (non-depressed group)	07/01/09	09/15/09									
Data entry	07/15/09	09/15/09									
Data analysis and thesis write-up	08/01/09	09/15/09									

3.6   Data Analysis

3.6.1   Internal consistency

To determine the extent to which items of the HDS-OA correlated with one another, a Cronbach's alpha coefficient was calculated for scores of the HDS-OA. It was anticipated that HDS-OA scores would have adequate internal consistency (alpha $\geq$ .80). Similar data analysis was conducted for the GDS and GDS-15, with the results of the GDS and GDS-15 compared to those of the HDS-OA.

3.6.2   Convergent validity

To demonstrate that the HDS-OA was measuring the construct that it was designed to measure, that of depression, scores were compared to an existing measure of

depression that possessed extensive validation evidence for the population of interest. In this research, convergent validity was assessed by calculating Pearson product-moment correlation coefficients (r) between the scores of the HDS-OA and those of the GDS and GDS-15. It was expected that HDS-OA scores would be statistically significantly and positively related to depression scores on the GDS and GDS-15, with correlations of > .80.

3.6.3 Discriminant validity

In order to examine the relationship between the HDS-OA and measures of theoretically different or less-related constructs to establish discriminant validity, this study used Pearson product-moment correlation coefficients (r) to examine the relationship between the scores of the HDS-OA and scores from the MMSE, BAI, and self-ratings of physical health. It was expected that cognitive status, health ratings, and anxiety scores would be less related to scores from the HDS-OA than the GDS and GDS-15. Scores of the GDS and GDS-15 were also correlated with scores of the MMSE, BAI, and self-ratings of physical health to compare values of discriminant validity between the HDS-OA and the GDS and GDS-15.

3.6.4 Criterion validity

Clinical diagnoses of depression and confirmation of the absence of a mood disorder in the community sample using the SCID-I/NP both served to provide the criterion measures in this study. In this research, differences between HDS-OA scores of the clinically depressed group and the non-depressed community group were used to

assess criterion validity. It was expected that HDS-OA scores would be significantly lower in the SCID-I/NP non-depressed community group than the clinically depressed group. Similar results were expected of the GDS and GDS-15. Results of the measures were compared.

As a further measure of criterion-related validity, and to determine the ability of the HDS-OA to accurately identify cases of depression from those who are not depressed, the Microsoft Excel statistical program Analyse-It was used to conduct a ROC curve analysis based on the scores for both groups. Based on the criterion measure determined by a depression diagnosis or a SCID-I/NP determined absence of depression, generated cut scores were evaluated for sensitivity, sensitivity, PPV, and NPV, and an optimum cut score was determined. The sensitivity, sensitivity, PPV, and NPV of the GDS and GDS-15 were also calculated for a range of cut scores and optimum cut scores were also determined for the GDS and GDS-15. Sensitivities, specificities, PPVs, and NPVs at the optimum GDS and GDS-15 cut scores were compared to these values for the HDS-OA. For each of the possible cut scores in the range, the ROC curve plots the proportion of true positives against the proportion of false positives (Swets et al,, 2000). The resulting AUC provided an overall indication of the accuracy of the measure. It was anticipated that the optimum cut score would result in an AUC of .80 or higher, demonstrating the usefulness of the HDS-OA as a case-finding depression screen. The AUC was also computed for the GDS and GDS-15 and compared to that of the HDS-OA.

3.7   Ethical Issues and Credibility

Before commencing the study, all participants were asked to sign an informed

consent form that outlined the nature of the study, procedures, time commitment,

responsibilities, potential risks, ability to withdraw, and statement of confidentiality (see

Appendices E and G). To protect confidentiality, identification (ID) numbers were

assigned to the results of each individual's questionnaire and interview results, and all

results were kept in a locked room at UBC. The master list with participant's names and

ID numbers are only available to Dr. Anita Hubley (principal investigator), Sherrie Myers

(co-investigator), and the Adult Development and Psychometrics Lab Coordinator, Lara

Russell. One limit to confidentiality was that the researchers were ethically obligated to

report any cases in which it was revealed that an individual was at imminent risk of harm

to themselves or others, and to solicit help for that individual.

An additional concern was related to the study's focus on experiences of negative

mood. In thinking about and discussing these experiences during the SCID-I/NP

interview, it was anticipated that some community sample individuals might experience

some distress during or after the session. Therefore, the community-based participants

were provided with information on community-based counselling services and resources

available to them in the greater Vancouver area as well as on-line (see Appendix E). This

information sheet also served an educational role. During any measure administration or

interview, if a participant indicated discomfort with a question, he/she had the right to

skip the question and the right to withdraw from the study at any time. If a depressed

group participant exhibited distress or disclosed self-harming intentions during measure

administration, study protocol dictated that the session would be concluded and the

treating physician or psychologist would be immediately contacted, although this did not occur in any session with a participant.

In addition, participants were not provided with their individual results on either the measures or the SCID-I/NP interview, but a summary of the overall study results would be sent to interested participants when the research is completed. Individuals had the opportunity to identify their interest and provide their mailing address or email address on the informed consent sheet. The consent form was not assigned an ID number in order to maintain confidentiality.

To ensure credibility, the study was based out of the Department of Educational and Counselling Psychology, and Special Education in the UBC of Faculty of Education, and the research was overseen by Dr. Anita Hubley, a faculty member with research experience and expertise in the measurement of depression in older adults. Faculty supervision also provided the researcher with access to the SCID-I/NP and supervision of its administration.

The University of British Columbia-Providence Health Care Behavioral Research Ethics Board provided ethical approval for this study. This approval allowed the research to be conducted with non-depressed participants from the community, but a further certificate of approval was provided by Providence Health Care for the research to be conducted with depressed individuals at Mount Saint Joseph's and St. Paul's Hospitals. In addition, the Vancouver Coastal Health Research Institute provided further approval to conduct recruitment and testing of depressed individual at Vancouver General Hospital and within the North Shore Community service delivery area of Vancouver Coastal Health. See Appendix G for copies of all ethical certificates and approvals.

3.8   Funding

This research was awarded a grant from the British Columbia Network for Aging Research (BCNAR) for data collection purposes. Funds were approved and used to purchase measures (e.g., MMSE, BAI), cover administrative costs of photocopying, pay for travel expenses, and pay the salary of research assistants who assisted with the recruitment of, and the administration of measures to, the non-depressed group. The University of British Columbia's Faculty of Education also funded this research. These funds were used to provide stipends to research participants to gratefully acknowledge the time they spent participating in the study.

Chapter 4 - Results

4.1   Sample Description

The sample consisted of 18 clinically diagnosed depressed subjects being treated for varying levels of depression at the time of test administration and 18 community-based non-depressed subjects matched on age and gender variables to the depressed group participants, for a total of 36 participants. Although the depressed group consisted of 18 individuals, this group included 16 inpatients and outpatients being treated for depression at the time of measure administration who were recruited from four hospitals and one community health center, and it also included two community-based individuals who were diagnosed by the SCID-I/NP with current depression and subsequently added to the depressed group. In addition, all participants in the study scored 21 or greater on the MMSE. Therefore, no individual's results were excluded from the study due to moderate or severe cognitive impairment. Table 4.1 provides a demographics summary.

Table 4.1 Demographics Summary

| | Group Demographics | |
Demographic Variable	Depressed (%) N=18	Non-depressed (%) N=18
Gender		
Male	22.2	22.2
Female	77.8	77.8
Age		
Age Range (yrs)	67 to 86	68 to 88
Mean Age (yrs)	78.22	78.22
Standard Deviation (yrs)	6.45	6.30

Demographic Variable	Group Demographics	
	Depressed (%) N=18	Non-depressed (%) N=18
Ethnicity		
Caucasian	83.3	94.4
South Asian	11.1	0
Métis	5.6	5.6
Marital Status		
Widowed	55.6	44.4
Married	22.2	27.8
Separated/Divorced	16.7	16.7
Single/Never married	5.6	11.1
Highest Education Level Achieved		
Doctoral Degree	0	11.1
Masters Degree	11.1	22.2
Bachelors Degree	0	27.8
Some University	33.3	11.1
Diploma	16.7	16.7
High School	5.6	5.6
Some High School	11.1	5.6
Elementary School	22.2	0
Self-rated Health		
Median Score	3	4
Mean Score	3.22	3.94
Standard Deviation	1.11	0.94
Anti-depressant Usage		
Yes	83.3	16.7
No	16.7	83.3

It is of note that 16.7% of individuals in the non-depressed group were taking anti-depressant medication at the time of measure administration. These subjects confirmed that the medication was being taken in low dosages as a sleep aid and not being used to treat a current depression. The SCID-I/NP also confirmed that these individuals were not depressed at the time of measure administration.

It is also worthy to highlight the differences in the demographic variables between groups with respect to marital status, education, and self-rated health. Given the observed differences, it was of interest to conduct a chi-squared test for independence to assess if there were statistically significant differences between groups based on marital status and education levels, but this analysis could not be conducted given that it was not possible for 80% of the cells in this analysis to contain a minimum of 5 subjects as a result of the limited sample size. The levels of marital status could not be combined in a meaningful way to conduct the analysis, and even when the education levels were grouped into only three levels, the minimum cell size assumption was still not met. A Mann-Whitney U test was conducted to test for a significant difference between the median self-rated health scores of both groups. The self-rated health scores ranged from 1 to 5, with higher scores indicating higher levels of self-rated health. The Mann-Whitney U test revealed no significant difference in the self-rated health levels of the depressed (Md = 3, n = 18) and non-depressed (Md = 4, n = 18) groups, U = 103.5, z = -1.92, p = .06, r = .32.

4.2   Internal Consistency

Internal consistency reliability of HDS-OA scores, using Cronbach's alpha, was

.90. Cronbach's alphas for the GDS and GDS-15 scores were .94 and .89, respectively.

4.3   Convergent Validity

The relationship between the scores on the HDS-OA and GDS was positive,

strong, and statistically significant with a correlation of .89. Similarly, the correlation of

scores between the HDS-OA and GDS-15 was positive, strong, and statistically

significant at .88.

4.4   Discriminant Validity

See Table 4.2 for a summary of convergent and discriminant measure correlations

for the HDS-OA, GDS, and GDS-15. All of these correlations were statistically reliable

at the .05 level, with the majority being statistically significant at the .01 level.

Table 4.2 Correlations Summary

Measure	BAI[a]	MMSE	Self-rated Health	GDS-15	GDS
HDS-OA	.74 (.000)	-.43 (.009)	-.49 (.002)	.88 (.000)	.89 (.000)
GDS	.79 (.000)	-.37 (.026)	-.57 (.000)	.97 (.000)	
GDS-15	.77 (.000)	-.42 (.011)	-.59 (.000)		

Note. Level of two-tailed statistical significant is denoted in brackets () below each
correlation.

[a]Alpha for the BAI was .90

Steiger's Z-test for correlated correlations within a population, using the freely

available statistical program FZT Computator (downloadable at

http://psych.unl.edu/psycrs/statpage/regression.html), was used to investigate if there

were any statistically significant differences between the correlations of the HDS-OA

with the discriminant measures (i.e., BAI, MMSE, and self-rated health item) and those

of the GDS with the same discriminant measures. The comparison between the HDS-OA

and GDS correlations with the BAI, MMSE, and the self-rated health item resulted in $Z =$

1.14, $Z = 0.77$, and $Z = 1.20$, respectively. Similarly, the comparison between the HDS-

OA and GDS-15 correlations with the BAI, MMSE, and the self-rated health item

resulted in $Z = 0.68$, $Z = 0.08$, and $Z = 1.31$, respectively. A significance level of $p < .05$

for a two-tailed test is achieved only if $Z > 1.96$. Therefore, none of these correlations

showed any statistically significant differences from one another.

4.5 Criterion Validity

4.5.1 Group differences

Differences in HDS-OA scores between depressed and non-depressed groups

were examined using the non-parametric Wilcoxon Signed Rank Test because the paired

samples parametric test assumption of normality of scores was violated for the non-

depressed group. A large statistically significant difference was found between the

depressed (median = 7.0) and non-depressed (median = 0.0) group scores for the HDS-

OA, $z = -3.63$, $p < .001$, r = .61. A large statistically significant difference was also

revealed between the depressed (median = 17.0) and non-depressed (median = 2.0) group

scores for the GDS, $z = -3.63$, $p < .001$, r = .61. Similarly, a large statistically significant

difference was found between the depressed (median = 7.5) and non-depressed (median = 1.0) group scores for the GDS-15 items, $z = -3.63$, $p < .001$, $r = .61$.

4.5.2 AUC, sensitivity, specificity, PPV, NPV

The AUC for the HDS-OA was .96 ($p < .001$). A cut score of 4 yielded an optimal balance between sensitivity (94.4%), specificity (88.9%), PPV (89.5%), and NPV (94.1%), with a greater emphasis placed on selecting a cut score that identified depressed individuals (i.e. sensitivity). See Table 4.4 for the sensitivity, specificity, PPV, and NPV for a range of HDS-OA cut scores.

Table 4.3 Sensitivity, Specificity, PPV, and NPV for HDS-OA Cut Scores

HDS-OA Cut-Score[a]	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0	100	0	50.0	--
1	94.4	61.1	70.8	91.7
3	94.4	83.3	85.0	93.8
4	94.4	88.9	89.5	94.1
5	88.9	100	100	90.0
6	77.8	100	100	81.8
7	55.6	100	100	69.2
8	44.4	100	100	64.3
10	33.3	100	100	60.0
11	16.7	100	100	54.5
14	11.1	100	100	52.9
15	5.6	100	100	51.4

[a]Maximum score on the HDS-OA is 16

The AUC for the GDS was .93 ($p < .001$). For this measure, a cut score of 7 or higher yielded an optimal balance between sensitivity (88.9%), specificity (88.9%), PPV (88.9%), and NPV (88.9%), again with a greater emphasis placed on selecting a cut score that identified depressed individuals (i.e. sensitivity). See Table 4.5 for sensitivity, specificity, PPV, and NPV for a range of GDS cut scores.

Table 4.4 Sensitivity, Specificity, PPV, and NPV for GDS Cut Scores

GDS Cut-Score[a]	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0	100	0	50.0	--
1	94.4	22.2	54.8	80.0
2	94.4	38.9	60.7	87.5
3	94.4	55.6	68.0	90.9
4	94.4	66.7	73.9	92.3
6	88.9	77.8	80.0	87.5
7	88.9	88.9	88.9	88.9
10	83.3	94.4	93.8	85.0
11	83.3	100	100	85.7
12	77.8	100	100	81.8
14	72.2	100	100	78.3
15	61.1	100	100	72.0
16	55.6	100	100	69.2

[a]Maximum score on the GDS is 30

The AUC for the GDS-15 was .93 ($p < .001$). In this case, a cut score of 3 or higher yielded an optimal balance, in the context of this study, between sensitivity (88.9%), specificity (83.3%), PPV (84.2%), and NPV (88.2%). See Table 4.6 for sensitivity, specificity, PPV, and NPV for a range of GDS-15 cut scores.

Table 4.5 Sensitivity, Specificity, PPV, and NPV for GDS-15 Cut Scores

GDS-15 Cut-Score[a]	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
0	100	0	50.0	--
1	94.4	38.9	60.7	87.5
2	88.9	72.2	76.2	86.7
3	88.9	83.3	84.2	88.2
4	83.3	100	100	85.7
5	77.8	100	100	81.8
6	66.7	100	100	75.0
7	55.6	100	100	69.2
8	50.0	100	100	66.7

[a]Maximum score on the GDS-15 is 15

Chapter 5 - Discussion

The purpose of this present study was to examine the psychometric properties of the HDS-OA using a sample of depressed and non-depressed older adults in order to provide reliability and validity evidence to support the inferences made from its total scores and support the use of this measure as an accurate screen for depression in older adults. In addition, the psychometric properties of the GDS and GDS-15 were investigated with the goal of determining if the HDS-OA is as useful as measure as these well known and well used geriatric depression screens. The following is a review and discussion of the findings of the study from a psychometric perspective, as well as a discussion of the strengths and limitations of the study, what was learned from a personal perspective about depression and its measurement in older adults, and the lessons learned in conducting research with older adults.

5.1   Discussion of Numerical Results

The alphas of .90 and .94 for the overall sample for HDS-OA and GDS total scores, respectively, are both excellent. The alpha reported for the HDS-OA is slightly lower than the value of .94 found in preliminary results by Hubley et al. (2009), but the alpha found for the GDS corresponds to values reported in previous research with older adults (Dunn & Sacco, 1989; Low & Hubley, 2006; Yesavage et al., 1983). Internal consistency reliability of the GDS-15 items was .89, similar to that for the equivalent length HDS-OA. This value is higher than alphas reported for the GDS-15 in previous research conducted with samples of depressed older adults (Almeida & Almeida, 1999; van Marwijk, Wallace, De Bock, Hernans, Kptein, & Mulder,1995).

The statistically significant, positive, and strong correlation of .89 between the HDS-OA and GDS and the similar correlation of .88 between the HDS-OA and GDS-15, supports the convergent and, therefore, the construct validity of the HDS-OA. These values provide evidence that the HDS-OA is highly correlated to other validated scales that measure the same construct in the same population. Similar correlations were found for the GDS and GDS-15 with other depression measures in previous research (Almeida & Almeida, 1999; Dunn & Saaco, 1989; Yesavage et al., 1983). The higher the correlation, the more the measures are similar. Therefore, in this case, the correlations indicate that there is enough of a difference to suggest that the scales are measuring a similar construct, but not so similar that the HDS-OA is an entirely redundant measure.

The lower statistically significant and negative correlations of -.43 and -.49 found between the HDS-OA and measures of different constructs (i.e., cognitive functioning measured by the MMSE and an individual's perception of their current level of overall health measured by the self-rated health item) indicate that the HDS-OA is less related to these measures as it is to the GDS and GDS-15, measures of the same construct. However, it is noteworthy that these correlations reveal that cognitive status and self-perceived level of health account for up to 18.5% and 24.0% of the variance in HDS-OA scores, respectively. These lower correlations though, as compared to that with the convergent measure, provide support for the discriminant validity of the HDS-OA. The results are in-line with discriminant validity evidence found for other depression measures (Osman et al., 2004; Segal et al., 2008; Snyder et al., 2000).

When these discriminant correlations for the HDS-OA were compared to the same correlations for the GDS (-.37 and -.57 for the MMSE scores and self-rated health

item, respectively), cognitive status accounted for a lower variance in GDS scores at 13.6% and self-perceived health level accounted for a greater variance in the GDS scores at 32.5%. As well, when compared to the same correlations for the GDS-15 (-.42 and -.59 for the MMSE scores and self-rated health item, respectively), cognitive status accounted for a similar variance in GDS-15 scores at 17.6% and self-perceived health level accounted for a greater variance in GDS-15 scores at 34.8%. However, Steiger's Z-test found no statistically significant differences between the discriminant correlations found for the HDS-OA, GDS, and GDS-15 with either the MMSE or self-rated health item. These results suggest that, with this sample of older adults, the HDS-OA did not demonstrate a significantly higher level of discriminant validity as compared to the GDS or GDS-15 with respect to the MMSE and the self-rated health item.

The statistically significant and positive correlation of scores on the HDS-OA with those of the BAI was found to be higher than the other discriminant measures at .74, but not as high as the convergent validity coefficient of .90 between the HDS-OA and GDS or the correlation of .88 between the HDS-OA and the GDS-15. The correlations of scores on the GDS and GDS-15 with those of the BAI (.79 and .77, respectively) were also higher than with the other discriminant measures. With these values being so close in magnitude though, Steiger's-Z found no statistically significant difference between the correlation of scores on the HDS-OA with scores on the BAI as compared to the correlations of scores on the GDS and GDS-15 with scores on the BAI. Due to the noted overlap in symptomatology between depression and anxiety in older adults, the use of the BAI as a discriminant measure did not produce as low a correlation with the HDS-OA,

GDS, or GDS-15 as the MMSE or self-rated health item. These results still provide support for the discriminant validity of the HDS-OA.

Including the BAI as a discriminant measure in this study because it measures a different, yet somewhat related, construct that shares some overlapping symptomology, potentially sheds new light on what may be considered a discriminant measure. This correlation reveals that indeed the HDS-OA and BAI are measuring some of the same symptomatology, but the correlation is different enough from that of the convergent measure to say that the BAI is also measuring something separate. In fact, these results also potentially shed a different light on the concept of the dichotomy of discriminant and convergent validity. The idea of discriminant and convergent validity would be better viewed on a spectrum from a measure being fully divergent to fully convergent, but with an area somewhere in between where a measure could support evidence of both convergent and discriminant validity to varying degrees. The dichotomous nature of discriminant and convergent validity may not be as black or white as the word-pair itself infers. It could be said that a correlation with a scale that measures a somewhat different construct, which is slightly lower than a correlation with a scale measuring the same construct, reveals evidence of some convergent but also some evidence of discriminant validity. The slightly lower correlation infers that, to some degree, a different construct is being measured. Therefore, the provision of evidence of discriminant validity could be based on that of relativity. In other words, key questions that could be asked to determine if some level of discriminant validity evidence is provided might be (a) "Is the correlation higher or lower than that provided by a comparison to a validated measure of the same construct?", and (b) "Is the correlation similar to correlations found for other known

discriminant measures?". The inclusion of a measure such as the BAI in this study is also

a more stringent test of the discriminant validity of the HDS-OA, as there is overlapping

symptomatology between depression and anxiety. In this case, the BAI has provided

additional support for the discriminant validity of the HDS-OA. Moreover, the HDS-OA

has proven itself to be a match to (or even slightly better than) the GDS and GDS-15 in

its ability to discern depression from anxiety.

With respect to the evaluation of criterion validity, the finding that the mean

HDS-OA group scores were significantly different between the non-depressed and

depressed participants provides criterion-related evidence to support the ability of the

HDS-OA to differentiate individuals with depression from those without. This evidence,

therefore, provides support for the use of the HDS-OA total scores to screen for

depression in older adults. The results for the GDS and the GDS-15 were similar,

indicating that the HDS-OA is an equally good measure, based on this analysis, as

compared to the GDS and GDS-15.

With further respect to criterion-related evidence provided by the present study,

the ROC curve analysis revealed a statistically significant AUC of .96 for the HDS-OA,

which indicates that it is an excellent case-finding screen. Although the AUC for the

GDS was slightly lower at .93, no statistically significant difference was found between

these values ($p = .19$). The AUC for the GDS-15 was .93. Similar to the GDS, no

statistically significant difference was found between the AUC values for the HDS-OA

and GDS-15 ($p = .23$). Therefore, these results indicate that the HDS-OA, GDS, and

GDS-15 are all useful measures in identifying truly depressed from truly non-depressed

older adults.

The optimum cut score for the HDS-OA that achieved the optimum balance between sensitivity (94.4%), sensitivity (88.9%), PPV (89.5%), and NPV (94.1%) was a total score of 4 or greater. In selecting a cut score, the consequential validity of the measure was considered. The consequence of not identifying truly depressed individuals must be balanced against the consequences of a positive screen for individuals who are later diagnostically confirmed not to be depressed. The repercussions of the latter might include the stigma that is attached to a diagnosis of depression, the resulting concern potentially experienced by these individuals, and the costs related to subsequent diagnostic testing (Swets et al., 2000). However, considering the potential emotional and physical costs to depressed older adults and the trickle-down financial costs that geriatric depression bears on the health-care system for those who go undiagnosed as a result of the use of an inappropriate cut score (Callahan et al., 1994; Simon et al., 1995), the emphasis in the present research was placed more on sensitivity than specificity when selecting appropriate cut scores for this population. In support of this approach, Almeida and Almeida (1999) suggested that, when selecting cut scores for the GDS, a cut score that favors sensitivity and NPV should be selected, since the usefulness of a scale depends on its ability to accurately identify cases of depression.

For the GDS-15, a cut score of 3 or higher yielded an optimal balance between sensitivity (88.9%), specificity (83.3%), PPV (84.2%), and NPV (88.2%), again with a greater emphasis placed on identifying depressed individuals (i.e., sensitivity). These results differed slightly from the ROC curve analysis results published for the GDS-15 by Almeida and Almeida (1999). The authors reported that the use of the cut-off point 4/5 for the GDS-15 produced a sensitivity, specificity, PPV, and NPV of 97.0%, 54.8%,

69.6% and 94.4%, respectively, when DSM-IV diagnostic criteria for a major depressive episode were used as the "gold standard". Similar to the present research, an optimum cut score of 2/3 was reported by van Marwijk et al. (1995) with a non-psychiatric sample from a general medical practice. The sensitivity, specificity, PPV, and NPV values found in this research for the GDS-15 were slightly lower than those found at the optimum cut score for the HDS-OA, indicating that the HDS-OA did a slightly better job with this sample of correctly identifying more cases of true depression and recording fewer false positives as compared to the GDS-15.

For the GDS, a cut score of 7 or higher yielded an optimal balance between sensitivity (88.9%), specificity (88.9%), PPV (88.9%), and NPV (88.9%), with a greater emphasis placed on selecting cut score with high sensitivity. These results suggest that the GDS and GDS-15 performed similarly. However, these values were lower than those found at the optimum cut score for the HDS-OA, with the exception of specificity, indicating that the HDS-OA did a slightly better job with this sample of correctly identifying more cases of true depression as compared to the GDS.

The cut score of 7 found for the GDS in this study appeared to be a relatively low cut score for a 30-item measure. At the suggested cut scores of 10 (Rapp et al., 1988) and 11 (Brink et al., 1982) for the GDS, the sensitivity was found to be 83.3% for both cut scores and specificity was 94.4% and 100%, respectively. If these suggested values were used to screen the individuals in this sample, several who were being treated for depression at the time of measure administration would have been excluded from being positively screened. The low optimum cut score found for the GDS prompted an investigation into whether the depressed group mean scores for the GDS in this study

were lower than those found in similar studies. Indeed, Yesavage et al. (1983) reported a mean score of 22.85 on the GDS for a severely depressed group, but also a lower mean score of 15.02 for a mildly depressed group, the latter of which was similar to the GDS mean score of 15.83 for the depressed group in this study. As well, the HDS-OA mean score for depressed participants in this study was 1.87 points lower and the median score was 3 points lower than preliminary data reported by Hubley et al. (2009). The lower mean score for the HDS-OA and lower mean score for the GDS, possibly reflecting overall milder depressive symptomatology in this sample, was unanticipated given the clinical diagnosis of the depressed individuals.

One of the reasons for the lower mean depression scores, and perhaps the lower optimum cut score obtained for the GDS, could be that some of the depressed participants, although actively being treated for depression, were at different stages of treatment and may not have been experiencing a severe MDE at the time of testing. Since testing did not occur immediately before patients were treated for depression due to the difficulty of such timing and recruitment, some individuals could have been in partial remission from a MDE and experiencing fewer symptoms. As well, given that the physician's diagnosis did not specify the severity of the depression in the majority of cases, some subjects could also have been experiencing another milder form of depressive disorder, such as minor depressive disorder or dysthymia (which was only specifically identified in two cases). It may have been the case that the depressed individuals who agreed to participate in the study were those who were feeling well enough to do so (i.e., possessing more energy and interest than individuals who were experiencing severe depressive symptomatology). Related to this, it was my experience

that the health care providers who assisted in the recruitment of depressed individuals carefully screened potential participants and invited only those individuals who they felt were mentally well enough to participate to take part in the study. Marwijk et al.'s cut score results of 2/3 for the GDS-15 appear to be in-line with the results found here, most likely as a result of the use of a similar sample, which also did not include psychiatric in-patient older adults, who possibly experienced less severe depressive symptomatology as a whole. However, it is also possible that lower scores could be attributed to a level of depressive symptom underreporting by some of the depressed subjects in this study, as has been suggested in previous research on depression screening in older adults (Brodaty, Thompson, & Mitchell, 2005; Davidson, McCabe, & Mellor, 2009; Lyness, Cox, & Curry, 1995). This could be especially possible in the case of a study such as this whereby an unfamiliar researcher, who was not part of the medical system in which subjects were being treated, was conducting the test session with little time to build rapport.

5.2   Comparison of the HDS-OA to the GDS and GDS-15

Based on the above discussion of the reliability and validity evidence for the HDS-OA, GDS, and GDS-15, it appears that the HDS-OA is relatively equal to the GDS and GDS-15 in terms of its psychometric properties. The internal consistency of the HDS-OA scores was slightly lower than that found for the GDS but similar to the GDS-15, although all measures showed excellent internal consistency. The comparison of the correlations of each depression measure with the discriminant measures also showed no statistically significant differences. Therefore, the HDS-OA showed neither more nor less

discriminant validity than the GDS or GDS-15. In addition, the HDS-OA, GDS, and GDS-15 revealed similar results for the between-groups mean score comparisons. So, it can be said that the HDS-OA performs as well as the GDS and GDS-15 in differentiating the scores of depressed and non-depressed groups. In terms of the ROC analysis, the HDS-OA showed a slightly larger AUC than the GDS and GDS-15, indicating that it may be a slightly better case-finding screen, but there was no statistically significant difference between these values.

With respect to slight psychometric differences that were found between the measures, the HDS-OA revealed a slightly higher specificity, PPV, and NPV as compared to the GDS for the optimum cut scores identified, indicating that the HDS-OA was slightly better at identifying cases of true depression than the GDS with this sample. The HDS-OA also revealed a better sensitivity, specificity, PPV, and NPV at an optimum cut score as compared to the GDS-15, indicating that it performed slightly better than the GDS-15 in identifying cases of true depression and in recording fewer false positives. It is worthy to note that because the items of the GDS-15 were extracted from the GDS and not administered to research subjects as a separate measure, the findings from all the analyses for this measure are not irrefutably. However, the results do suggest comparable performance between the HDS-OA and GDS-15 and indicate that this is an area for future research and confirmation.

With the exception of the ROC curve analysis results for the HDS-OA at its optimum cut score, the results of the depression measures are very similar in many of the analyses performed. Therefore, mental and physical health care practitioners and researchers, the very individuals who may use the measure, may question if this evidence

is enough to support the use of the HDS-OA to screen for depression in older adults over the better-known and long-established GDS or the shorter GDS-15. There are a couple of design elements of the HDS-OA that may provide advantage over the GDS and GDS-15 that are worth highlighting here.

The HDS-OA includes a reminder of the two-week reference period at the beginning of each item but the GDS and GDS-15 do not contain such reminders. Therefore, it is less certain if subjects remember the one-week reference period for GDS or GDS-15 as they respond to the items of the questionnaire. This omission adds an extra element of uncertainty to the scores of these measures as compared to the HDS-OA.

Related to this, the two-week reference period of the HDS-OA is a longer period over which to identify if potential depression symptoms exist, as compared to the one-week timeframe for the GDS and GDS-15. Given that the intent of screening is to identify individuals who may benefit from diagnosis and subsequent treatment, the lengthened timeframe may provide a better window in which to identify the existence of symptoms of a time sensitive and variable construct such as depression and is in-line with the reference period used in the DSM-IV-TR. In the case of dysthymia for instance, whereby the major criterion, as defined by the DSM-IV-TR, is a depressed mood more days than not over the past two years (APA, 2000), it is possible that symptoms may not have occurred over the past week but may have occurred in the past two weeks. In fact, item 16 of the HDS-OA, "Over the past two weeks: Have you thought that the future looks hopeless?", was included in the measure to potentially identify the experience of dysthymia in individuals. Identifying individuals who may have had symptoms two

weeks ago, but not one week ago, is a more conservative and circumspect approach to screening for such a debilitating mental health issue.

In addition, given that the development period for depression can potentially be years in older adults versus days or weeks in younger adults (Berger et al., 1998; Huffstetler, 2001), a screen with a high sensitivity that is capable of better identifying cases of individuals with lower levels of depressive symptomatology, as may have been the case with some individuals in this study, appears to be useful. It would be prudent to identify these individuals before a mild depression possibly progresses to a full-blown MDE. The use of a screen such as the HDS-OA could result in the reduction of health care costs through the recognition and subsequent treatment of depression at an earlier stage.

As well, the HDS-OA, in contrast to the GDS and GDS-15, includes several somatic items to identify depressive symptoms in individuals who might manifest depression in this way. In doing so, it provides individuals with an opportunity to express their depressive symptoms in a way that is consistent with their experience of depression, but also in a socially acceptable way and/or in a manner that is culturally acceptable for them (Lewinsohn et al., 1997). Waza, Graham, Zyzanski, and Inoue (1999) reported that Japanese family physicians recorded more somatic complaints such as headaches, abdominal distress, and neck pain for those with depression than physicians in the United States and noted that these symptoms had a strong cultural significance for the Japanese patients. It has also been reported that men, due to social expectations and conditioning, express depression in somatic ways more so than women, who often express depressive symptoms in more emotional ways (Brownhill, Wilhelm, Barclay, Parker, 2002; Oliver &

Toner, 1990). As well, Kupfer, Kuhl, and Regier (2009), in an editorial on improving diagnostic systems for later life mental health issues, noted that older adults in general might be less likely to endorse affective symptoms. Davidson et al. (2009) reported that depressed mood was the most often omitted symptom for those who were diagnosed with depression assessed through the use of a depression screen, a diagnostic interview, and the help of an informant interview. Unlike the GDS and GDS-15, the HDS-OA provides a balance of questions based on the somatic, affective, and cognitive symptoms of depression in order to identify symptoms that may be expressed in any of these ways.

Although doubt exists as to the usefulness of somatic items in the assessment of depression in older adults due to their confound with physical illness (Yesavage, 1983; Balsis & Cully, 2008; Brink et al., 1982), Norris, Snow-Turek, and Blankenship (1995) found that low energy, strained effort, and sleep disturbances contributed to the assessment of depression in older adults, though appetite changes and level of sexual activity did not. The authors reported that institutionalized older adults expressed more somatic symptoms of depression as compared to those living in the community. Norris et al. noted that clarification regarding the issue of inclusion or exclusion of somatic items when assessing geriatric depression rests in building awareness of group differences in depressive symptomatology and in evaluating the helpfulness of specific somatic symptoms.

Even though the somatic symptoms of a physical illness and the somatic symptoms of depression can appear quite similar and confound the screening process, true somatic symptoms of depression that are not due to a medical condition or substance use are usually linked to events or circumstances in an individual's life (Gatz, 2000; Gatz

& Fiske, 2003; Kraaij & Garnefski, 2002; Vink et al., 2007). If there is doubt as to the etiology of the symptom (i.e., depression versus a physical ailment), a physical origin of a somatic symptom that is flagged during a depression screening may be ruled out or confirmed through medical testing that often occurs concurrently with psychiatric or psychological testing of older adults who are under medical care. If ruled out, a psychogenic origin can be further investigated through subsequent diagnostic testing for depression. As Stickle and Ondera (2006) suggested, in addition to screening for the depressive symptoms specifically seen in older adults, a more comprehensive assessment should also be conducted into how the distress is expressed as well as the impact of physical illness on depressive symptoms.

Finally, to further support the use of the HDS-OA, it could also be said that, for those professionals in research, counselling, or health care who are interested in using a depression screen that is based on DSM criteria for depression, the HDS-OA is a better choice than the GDS or GDS-15. The items of the HDS-OA are grounded in the current DSM criteria for depression, which is often used as a "gold standard" measure by which to diagnose depression in all persons (Almeida & Almeida, 1999; Shear et al., 2000; Steiner et al., 1995). Two additional questions on the HDS-OA about medication usage and bereavement, in accordance with the DSM-IV-TR substance use and bereavement specifiers for depression, may also help to highlight if a potential depression is linked to the use of a medication or to the grieving process. Until updated depression criteria are provided for older adults, the current DSM symptomatology for depression is an accepted standard by which individuals are assessed. With respect to the ideal of standardization in measurement, there is merit in a scale that is developed based on a known standard.

Despite this ideal, it is important to note that the current DSM criteria have been questioned regarding their ability to accurately measure depression in later life. In fact, Davidson et al. (2009) reported that individual diagnostic interviews based on DSM depression criteria and self-report questionnaires may be insufficient to detect depression in the elderly due to an underreporting of symptoms in this population, and suggested the addition of an informant interview for accurate diagnosis. With respect to specific DSM diagnostic criteria, Balsis and Cully (2008) suggested that somatic as well as cognitive symptoms might be less discriminating in older adults as compared to younger adults. The authors reported that older adults were indeed more likely than younger adults to endorse somatic items versus affective or cognitive items, but the authors attributed this to an increase in sleep problems, weight change, and health issues as people age, not just with depression. However, the authors did note that when depression levels were held constant across young and older adults and when DSM items were considered together, the DSM criteria for depression as a whole worked equally well across age groups.

Because there are questions about the accuracy of the current DSM criteria with older adults (Kupfer et al., 2009), a report by the task force that is contributing to the development of the DSM-V, which expected to be released in May 2012 (APA, 2009), indicated that the lifespan development group is examining incorporating age-related subtypes into the diagnostic criteria of text revisions of the DSM-V (APA, 2008). In this same report, the mood disorders work group discussed needed changes to the DSM-IV-TR criteria for a major depressive episode related to the addition or removal of symptoms, sub-threshold depression and suicide risk, as well as the bereavement specifier. Although a subgroup of the mood disorders work group has been looking into

possible differences in depression symptoms across gender and cultural groups, this work group did not report on any proposed changes to the diagnostic criteria for depression specifically with respect to older adults (APA, 2008). It has been indicated that decisions on how to incorporate the issues of older adults into the DSM-V "remain at a distance" (Kupfer et al., 2009, p. 358.). Although the awareness of these diagnostic limitations exists, implementation is lagging behind knowledge. At this time, the HDS-OA, which is based on the current DSM criteria for depression, may offer the best compromise between the affective, somatic, and cognitive symptoms of depression that older adults have been documented to experience and report.

Finally, in favor of the HDS-OA, it could also be argued that the 16-item HDS-OA is significantly shorter than the original 30-item GDS. This is indeed the case but, as discussed, the GDS-15 has also shown to be a reliable and good screening instrument for major depression in the elderly, with its total score showing to be a reliable measure of the severity of the depression (Almeida & Almeida, 1999). Given the existence of a similar length depression screen for older adults, the GDS-15, a comparison to the results of the HDS-OA and GDS-15 in this present study was conducted. As discussed, this comparison revealed similar psychometric evidence between the two measures. Although the length of the HDS-OA is neither an advantage nor a disadvantage when compared to this shorter-form version of the GDS, it is suggested that the HDS-OA performed at least as equally as well as the similar length GDS-15 with this sample of older adults.

5.3   Study Strengths and Limitations

The strengths of this present study included the use of a diagnostic measure, the SCID-I/NP, to confirm the absence of depression in the community sample, inclusion of a clinically diagnosed depressed group of older adults, use of experimental matching of age and gender, use of a double-blind procedure, counter-balancing of measures, and inclusion of a cognitive status exam in order to identify moderate to severe cognitive impairment in participants.

The use of the SCID-I/NP to diagnostically confirm that subjects in the non-depressed group were in fact not depressed added confidence in the scores of the depression measures of the non-depressed group. It also acted as a criterion measure enabling a comparison between groups. In addition, the use of the SCID-I/NP enabled a depression diagnosis to be confirmed for two individuals recruited for the non-depressed group who were in fact suffering from depression. This SCID-I/NP depression diagnosis acted as the criterion measure for the results of these two subjects when they were added to the depressed group.

With respect to the inclusion of a clinically diagnosed depressed group, it was desired that all depressed subjects have a DSM-IV-TR based diagnosis of depression. Due to the number of hospitals and clinics participating in the study and the uncertainty of the basis of each physician's diagnosis of a patient, which was sometimes retrieved from medical records by other physicians or health care staff because the diagnosing physician was once-removed from the care facility, a DSM-IV-TR based diagnosis for all patients could not be confirmed. However, it is certain that all patients included in the sample did have a current diagnosis of depression (and not due to a general medical

condition or substance use) and were actively being treated for depression at the time of test administration.

The matching of gender and approximate ages between groups also added to the strength of the study by helping to control for group differences on the basis of these variables. Although age matching between the depressed and non-depressed groups was not exact and, therefore, allowed for less control over the age variable, it was anticipated that matching based on the use of a maximum two and a half year age difference younger or older would provide a sufficient level of control over this participant variable. Even though the age matching was conducted in wide five-year age intervals, the mean age of both groups was the same, adding more confidence in the level of control achieved over this variable.

With respect to administration procedures, for subjects in the non-depressed community based group, research assistants administered the depression, anxiety, and cognitive status measures so that I would be blind to the results of the questionnaires when I subsequently administered the diagnostic interview and they would be blind to my results. This approach limited bias in the results of the diagnostic interview and measures. Related to this, the administration of the GDS, MMSE, HDS-OA, and BAI to individuals in both groups was counter-balanced using a digram-balanced approach in order to limit fatigue effects due to order effects.

A final strength of the study was the use of the MMSE to confirm that all subjects were cognitively capable of participating in the study. The use of this measure reduced concerns that cognitive status could be a viable latent variable as its relationship to HDS-

OA scores could be examined and ensured that no scores were invalid due to severe cognitive impairment.

Although the study possessed several strengths, inferences made from the results are moderated by a number of obvious and potential limitations that were inherent in the design and scope of this study. These limitations relate to gender ratios, ethnic diversity of the sample, suspected group demographic differences, lack of a measure of social desirability, limited exclusion criteria, and sample size.

One obvious limitation relates to the gender ratio of participants. It would have been advantageous to have both groups comprised of the general population gender based ratio of approximately 50% men and 50% women, so that the results would have confidently been generalizable to both genders. The research was limited though by the higher ratio of women to men as age increases, the gender ratio of the clinically diagnosed group, and their willingness to participate in research. As noted though, exact gender matching was performed between both groups to control for gender variables.

In addition to gender inclusion limitations, the sample was largely composed of one ethnicity. The large majority of participants were of Caucasian ethnicity. Due to the requirement that participants speak, read, and write English fluently, many potential depressed subjects, particularly of East Indian and Asian decent, were excluded from the study due to the inability to read or speak English. Therefore, the results of this study can only be generalized to the sample included in this research, which was of a high percentage Caucasian ethnicity and fluent in the English language.

With respect to depressed and non-depressed group demographic differences, a Mann-Whitney U test revealed no significant difference in the self-rated health levels of the depressed and non-depressed groups, even though the median score of the depressed group was lower than the non-depressed group, indicating perceived lower levels of health. A chi-square test of independence could not confirm or disconfirm significant differences between the groups based on marital status and education variables, but there appeared to be notable differences between the groups, especially with respect to education levels and, to a lesser degree, with respect to the percentages of widowed subjects in each group. Although the suspected education level discrepancy between groups could be attributed directly to recruitment, it is a surprising finding given that subjects from both the depressed and non-depressed groups were recruited from the same catchment area. These observed differences in both marital status and education levels between the groups attenuate the results of the study because it could not be confirmed that the groups were equal on these variables. However, these findings also are strengths of the study in that they revealed unexpected potential differences between groups that might serve as a signpost for future research related to factors that inhibit or encourage the development of depression in individuals.

Another potential limitation of the study was based on what was not included in this research design and scope. In this research, no social desirability scale was administered to participants, so there was potential that subjects may have responded to questionnaire items and to diagnostic interview questions in a socially desirable manner that minimized depressive symptomatology, as has been suggested in previous research (Lewinsohn et al., 1997). This uncertainty is supported by research on depression in older

adults that has found an underreporting of depressive symptomatology (Brodaty et al., 2005; Davidson et al., 2009; Lyness et al., 1995). It might be expected that underreporting could be partially attributed to cognitive changes in the elderly that impair the reporting of symptoms but Teri and Wagner (1991), when they investigated the connection between dementia and reporting of depressive symptoms, could not find an association between levels of cognitive impairment and low reporting of symptoms of depression in older adults. Given some of the lower depression measure scores found in the present study, it is possible that an element of social desirability might have had an influence on the scores of this sample. This may have been the case with respect to the results of the mood questionnaires completed by one depressed participant in particular. This individual was actively being treated for depression at the time of test administration and hospital staff members were concerned for this individual's welfare, but the subject scored zero on both the GDS and HDS-OA. However, it is noted that although these results could be attributed to social desirability, they could also be related to the individual's denial of depressive symptoms, even though the subject was aware that she was taking part in a study because she had a diagnosis of depression.

Despite the concern regarding social desirability when screening for depression in older adults, the literature review revealed that few depression measure validation studies incorporated the use of a social desirability scale. As well, the addition of another scale would have lengthened the time commitment for participants to the research, which would have been especially challenging for individuals in the depressed group, who in many cases, were already challenged by both mental and physical health issues. Given these challenges, a scale of this type could not be included, but it may have shown

compelling results. Its inclusion may have helped to determine if, or to what degree, some of the scores found in this study were due to underreporting as opposed to suspected lower levels of depressive symptomatology.

As well, in terms of the design of the study, the use of anti-depressant medication was not deemed an exclusion criterion for the study since it would have severely limited the potential depressed subject pool. This is because individuals would have had to be asked to participate in the study at the outset of their treatment before medication was administered, thus adding a timing constraint and a recruitment burden. Patients who had not received any treatment for depression may have been less likely to participate in a study due to the often-inherent lack of energy and lack of interest that is part of depressive symptomatology. As well, it may have also been the case that patients who were referred to the hospitals or the health care centers participating in the study were already taking anti-depressant medication prescribed to them by their general physician. Therefore, from a research study feasibility perspective, it was also impractical for this reason to include anti-medication depression usage as an exclusion criterion. Given that most depressed participants were already receiving pharmaceutical treatment for depression, some of the depression scores found may not reflect the full extent of original depressive experiences, thus attenuating the efficiency of these screening tools in this research.

Finally, the relatively small sample size posed the risk of reducing power and increasing the risk of Type II error, but it was necessary to limit the study to 36 participants to maintain a reasonable timeline for study completion. This was due to the challenge of recruiting depressed participants through hospitals and health care centers

where staff members were required to recruit patients for the study on my behalf due to the need to maintain patient confidentiality. Despite the small sample size, the major analyses of the study all revealed statistical significance. As well, the number of statistical analyses required to evaluate the data, as specified in the research methodology, may have raised the chance of Type I error in the results, but again this concern may be tempered by the level of high statistical significance attained in many of the analyses.

5.4   Personal Observations About Depression and the Measurement of Depression in Older Adults

Although this is not a qualitative study, the experience of this research and the direct contact that I gained with both depressed and non-depressed older adults allowed me to gain a better understanding of the experience of depression in older adults. Many older adults who were depressed, as well as those who were not depressed but who had witnessed depression in their friends and/or loved ones, talked about the impact of loss, loneliness, isolation, lack of hope for a positive future, fear of what the future brings, fear of physical deterioration and incapacitation, the stigma of depression, and the lack of people in their lives who were sources of support. As a result, I believe that, in many cases of those who develop depression later in life, several of these elements combine at the same time to stress the coping resources of an individual who may already be challenged by physical illness, cognitive decline, and/or financial stressors, thus overcoming an individual and leading to the experience of depression. It seems that geriatric depression is often experienced as a state of having no hope for a personally

connected and fulfilling future comprised of feelings of joy, contentment, comfort, or pleasure. It appears to be a state of being that is marked by numbness and losses that are multiple and non-replenishable. As some of the literature suggested, depression in older adults appears to have roots in the absence of positive affect, rather than an increase in negative affect (Katz, 2000). This phenomenon could be summed up by the response of one depressed participants who noted that she was "no longer able to cry".

Due to the complexity of depression in older adults, as well as the great differences in the individuality of older adults due to the varied life experiences that each has had, the measurement of depression in this population may be as much an art as it is a science. For instance, it has been argued that the use of a dichotomous response format on a depression measure for older adults may be more useful than a multiple choice format due to general cognitive declines seen as individuals age (Olin et al., 1992; Yesavage et al., 1983), but many depressed and non-depressed subjects noted that they had difficulty selecting either a "yes" or "no" answer in response to some questions on the HDS-OA and GDS because their answer lay in between these extremes. These subjects indicated that it would have been helpful for them to have a scale from which to select a response that was more in the middle between "yes" and "no" (e.g. "not often", "sometimes", "often").

Individuals also commented about having difficulty with the interpretation of some of the language used in the GDS, such as the words "happy" and "exciting" because they felt that their emotional range was more restricted than it was in their younger years. Instead of feeling "happy", they mentioned that they feel "content" instead. Unfortunately, in answer to questions using words that denote an emotional extreme,

116

subjects who have difficulties such as this would have to interpret the question to create a better meaning for them and then answer appropriately. Conversely, they could read the question literally and answer "no" to any questions about feeling "happy" or experiencing life as "exciting", even though they may feel "content" with their lives, rather than "happy", or feel "pleased" by certain things in their lives, rather than feeling "excited". Neither feeling "content" nor feeling "pleased" is an emotional state reminiscent of depression though. Confusion such as this could lead to responses that add to a lack of confidence in the scores of depression screens that use emotional wording that is considered extreme by some.

As well, in relation to the use of language and its literal interpretation, it was also my experience during this research that many individuals read the literal meaning of the question and answered as such. For instance, several individuals had questions about two items on the HDS-OA that asked about "change" over the past two weeks. For example, item 2 on the HDS-OA asks a subject, "Over the past two weeks: Have your sleeping patterns changed?". The intent of the question is to determine if, over the past two weeks, the individual's sleeping pattern has been different than what is normal for them. If read literally, subjects would answer "yes" or "no" to whether a "change" had occurred in their sleeping pattern over the past two weeks. The concern is that a person may be experiencing depression but the change in sleeping pattern may have happened one month ago and is still different than it would normally be for them, but the actual change did not occur in the past two weeks. In this case, a change in sleeping pattern from what is normal for that person would not be identified as a symptom and the extent of depressive symptomatology could be underestimated. In cases where subjects asked

about these "change" questions, the intent of the item was explained by the measure administrator, but individuals were asked to answer the question as per their own reading of it.

The experience and the measurement of depression in older adults is indeed a complex matter but it is hoped that, through further research, areas of weaknesses and strengths can be identified and geriatric depression measures subsequently adapted and updated. In this case, these findings are anecdotal, but they do lead the way to areas of future research.

5.5   Research Lessons Learned

Through this work, I have also learned a number of lessons about conducting research involving older adults, especially those with physical and/or mild cognitive impairments. These lessons relate to the length of the test session, the length of the chosen measures, and the particular use of language on forms. These lessons are discussed below.

In future, I would ensure that every element of the test session is kept as short as possible, while still maintaining the integrity of the research. For instance, I would use as short a consent form as possible. The consent form used in this research was three pages long in order to include information required by the University of British Columbia-Providence Health Care Behavioral Research Ethics Board. In many cases with depressed group subjects, it took a large amount of time to explain the contents of the consent form and then also provide time for the subject to read though it. I believe

this often made the testing session longer than the 30 to 40 minutes anticipated and added an additional burden to the participant.

The inclusion of an anxiety measure in this research also added to the time required for test administration, but it was used to provide important information regarding evidence of discriminant validity and information as to the overlap of the symptomatology of anxiety and depression. The addition of a demographics form also added to the length of the test session. The form was only two pages long, but was yet another questionnaire for participants to complete. In future, it would still be important to retain a measure such as the BAI, when relevant to the study objectives, as well as collect demographic information. Although, one way to at least minimally reduce the time of the test session in future work would be to collect the demographic information on a condensed one sheet form so study subjects do not need to turn to a second page.

In selecting measures for research with older adults, I would ensure they were valid but as short as possible. For example, I would choose to administer the GDS-15 over the GDS, not only because all the hospitals and health care centers participating in the study were using the GDS-15, but also because the GDS is a long questionnaire to complete for someone who is experiencing depression and the GDS-15 has shown good reliability and validity evidence in past research (Almeida & Almeida, 1999; van Marwijk et al., 1995) and has shown similar psychometric properties as compared to the GDS in the present study. As well, if there were a shorter anxiety measure that provided the same level of psychometric evidence as the BAI with older adults, I would select it in lieu of the 21-item BAI in an effort to use measures that were as short as possible.

In addition to making decisions that shortened elements of the testing session where possible, I would also be cautious of the use of language on any forms that the subjects read. For example, the information letter given to potential depressed participants and the consent form for this group used the terms "depressed" and "depression" without consideration for the sensitivity that some individuals may have to these words and without realizing the stigma that some older adults have about depression. In fact, some individuals who had a diagnosis of depression were not asked to participate in the research by health providers because they were in denial of their depression and the health care staff did not want to aggravate the mental health of these individuals by asking them to participate in a study about depression in older adults. In future, I would instead use phrases such as "individuals who may have symptoms of depression" versus "individuals with depression" and "patient group" instead of "depressed group".

It was difficult prior to the beginning of the study to have been aware of, and have implemented, these above-discussed "best practices" for research involving older adults. These suggestions may each seem small and inconsequential on their own but, when combined, they may ensure a greater ease of test sessions for the subject and researcher, while still maintaining the integrity of the informed consent process and the integrity of the research. As it is with many things, it was only though experience that this knowledge was gained.

Chapter 6 - Conclusions

6.1   Data Analysis Conclusions

The study provides evidence supporting the reliability of scores and validity of inferences from the HDS-OA with a sample of depressed and non-depressed older adults. In this study, the HDS-OA, GDS, and GDS-15 all produced reliable (i.e., internally consistent) scores. The HDS-OA showed strong evidence of convergent validity with the GDS and GDS-15, and all measures showed equivalent levels of discriminant validity with the BAI, MMSE, and self-rated health item. As well, the HDS-OA, GDS, and GDS-15 each demonstrated the ability to differentiate those who were depressed from those who were not depressed. ROC curve analyses suggest that the HDS-OA, GDS, and GDS-15 are useful as case-finding screens for depression in older adults. The only point of difference between the measures was with respect the ROC curve analysis in which the HDS-OA showed slightly higher values for sensitivity, PPV, and NPV than the GDS and slight higher values for sensitivity, specificity, PPV, and NPV as compared to the GDS-15 at the optimum cut score for each measure. A cut score of 4 for the HDS-OA is suggested.

Based on the overall psychometric results, the HDS-OA is recommended at least as an equally viable geriatric depression screen as compared to the GDS with this sample of older adults. The results also suggest that the same is true when the HDS-OA is compared to the GDS-15. Given these findings, for clinicians or researchers who want to use a screen based on the DSM-IV-TR criteria for depression or for those who support

the inclusion of a balance of affective, cognitive, and somatic items in a geriatric depression screen, the HDS-OA is a better choice.

6.2   Areas for Future Research

With respect to future validation work that could be conducted for the HDS-OA, a test-retest study of the HDS-OA would be advantageous in order to measure test-retest reliability and reduce the threat of specificity of variables by adding more variability to test administration with respect to time of the day, duration, and location. Test-retest validity evidence could provide support for the generalizable of the results of the measure over a short period of time, given that the construct of depression can be transitory. It would also be advantageous to conduct a similar validation study of the HDS-OA comparing results to those found for the GDS-15, when the latter scale is not embedded in the original 30-item GDS, in order to potentially provide concrete evidence to support the use of the HDS-OA over an equivalent length depression screen. In addition, given the overlap of symptomatology between depression and anxiety, further studies could be conducted to determine the criterion validity of the HDS-OA with a diagnosed co-morbid depressed and anxious sample of individuals, since an additional diagnosis of anxiety or confirmation of the absence of an anxiety disorder was not included for either group in this study. As well, the addition of a psychometrically sound social desirability scale as part of the research package in future validation efforts could reveal interesting results that might be in accordance with the documented underreporting of depressive symptoms in older adults.

With respect to the doubt surrounding the applicability of the current DSM criteria for depression to older adults, in future, the HDS-OA may require updating and further validation work when consensus is reached on accurate diagnostic depressive symptomatology for older adults. These concerns may not be addressed in the DSM-V, which is to be released in spring of 2012, but ultimately they may be addressed in a DSM-V text revision given the current awareness of these issues by researches and medical professionals in the field of older adult mental health (APA, 2008; Kupfer et al., 2009).

This research has also informed potential future research on possible protective factors in depression in older adults. Based on the demographic differences observed between the depressed and non-depressed groups, the strength of the relationship between depression in older adults and marital status and education level variables could be further investigated. Although the present study was not designed to examine these relationships, the findings are important because they highlight future research that could be conducted on the nature of mitigating and aggravating factors in the development of depression in older adults.

As well, this research has highlighted areas for future research on the measurement of depression in older adults. As a result of some of the comments made by subjects around the restriction of their emotional range, work could be done in the area of the use of language and its impact on the measurement of mood in older adults. More research on the use of appropriate and accurate response formats for older adults could also be conducted, given the number of comments made by individuals regarding their difficulty in selecting between "yes" and "no" responses. However, this

suggestion is tempered by the fact that the present study did not include any individuals with moderate to severe cognitive impairment who may have especially benefited from the easy yes/no response format.

In addition, based on research regarding the underreporting of depressive symptoms in the elderly (Brodaty et al., 2005; Davidson et al., 2009; Lyness et al., 1995), and the lower likelihood of the endorsement of affective symptoms by some older adults (Kupfer et al., 2009), it may also be prudent to develop and validate an "other-report" older adult depression screen to be used by the family, friends, and/or care-givers of depressed older adults as an adjunct to the self-report HDS-OA. It may be wise for a measure such as this to be developed and validation research conducted after the release of diagnostic criteria that researchers and medical professionals agree better reflects the experience of depression in older adults.

6.3   Research Contributions and Counselling Implications

As outlined, depression in older adults often presents differently than depression in younger adults and is a prevalent issue that requires attention. Thus, there has never been a better time for new research that informs practice with this client group given that the population is quickly aging. The development and validation of new measures to screen for depression in the elderly is an important step towards ensuring that older adults who experience depression are identified and treated. The HDS-OA provides an easy to use, short, and accurate screen for the symptoms of depression based on current DSM criteria. Accurate screening will enable those individuals with a positive result to be diagnosed and treated as early as possible, and this is crucial given the debilitating and

detrimental effects that depression can have in the elderly in terms of lowered well-being, compounding of the effects of physical illness, increased suicide and mortality rates, and increased health care costs. Not only could the level of overall wellness in depressed patients be improved (Dixon, 2007), but also the costs for emergency room visits and overall medical treatment could be reduced (Callahan et al., 1994; Simon et al.1995).

The use of a screening tool such as the HDS-OA by counsellors or health care professionals must be weighed against the consequences that a positive screen or a subsequent positive diagnosis may have on the individual, such as potential labeling and stigmatization, which appear to be possible effects given the documented underreporting of depressive symptoms in older adults (Brodaty et al., 2005; Davidson et al., 2009; Lyness et al., 1995). If a person is falsely screened positive for depression, the impact is felt not just by the individual but also by the health care system in terms of the unneeded additional costs of diagnosis. Considering the impact that depression can have on the elderly though, the benefits of the use of a screening tool such as this appear to outweigh the risks.

The counselling of older adults is becoming increasingly important as the population becomes disproportionately older due to increasing life expectancies and the aging of baby boomers (Dixon, 2007). Counsellors can play a crucial role in the treatment of depression in older adults by offering prevention and treatment information to individuals and their caregivers, educating other helping professionals about the experience of depression in older adults, and being on the front lines of identifying the signs and symptoms of depression in older adults (Huffstetler, 2001). Through their own use of psychometrically sound depression screening tools with this

population and through the use of these proven measures by other mental health workers, primary care physicians, psychiatrists, and geriatricians to quickly and more accurately screen for depression in older adults, diagnosis can be confirmed and required treatment can be more quickly facilitated.

The use of a depression screen such as the HDS-OA in a counselling practice can also provide an avenue to further explore symptoms of depression and help to illicit dialogue about the specific experiences of a client with respect to a suspected depression. This dialogue can provide an outlet for, support for, and validation of the individual's experience. The existence and use of such a population-focused depression screen also acknowledges for the older adult client that depression is a significant issue that is worthy of treatment and it is not assumed just to be an acceptable and expected part of the aging process.

References

Albanese, M. A. (2003). Review of the Mini-Mental State Examination. In B. S. Plake, J. C. Impara, & R. A. Spies (Eds.), *The fifteenth mental measurements yearbook*. Retrieved September 5, 2008, from EBSCO Mental Measurements Yearbook database.

Almeida, O. P., & Almeida, S. A. (1999). Short versions of the Geriatric Depression Scale: A study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV. *International Journal of Geriatric Psychiatry, 14,* 858-865.

American Psychiatric Association. (2009). *DSM-V: The future manual.* Retrieved August 25, 2009 from http://www.psych.org/dsmv.asp.

American Psychiatric Association. (2008). *Current activities: Report of the DSM-V task force (November 2008).* Retrieved August 25, 2009 from http://www.psych.org/MainMenu/Research/DSMIV/DSMV/DSMRevisionActivit ies/DSMVTaskForceReport.aspx.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Baldwin, R. C., & Tomenson, B. (1995). Depression in later life: A comparison of symptoms and risk factors in early and late onset cases. *British Journal of Psychiatry, 167*, 649-652.

Balsis, S., & Cully, J. A. (2008). Comparing depression diagnostic symptoms across younger and older adults. *Aging and Mental Health, 12*, 800-806.

Beck A. T., Brown G., Steer R. A., Eidelson, J. I., & Riskind, J. H. (1987). Differentiating anxiety and depression: A test of the cognitive content-specific hypothesis. *Journal of Abnormal Psychology, 96,* 179-183.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting & Clinical Psychology, 56*, 893-897.

Beck, A.T., & Steer, R.A. (1993). *Beck Anxiety Inventory manual*. 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., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 53-63.

Beck, A. T., Weissman, A., Lester, D., & Trexler, R. (1974). The measurement of pessimism: The Beck Hopelessness Scale. *Journal of Consulting and clinical Psychology, 42,* 861-865.

Beck, J. G., Novy, D. M., Diefenbach, G. J., & Stanley, M. A. (2003). Differentiating anxiety and depression in older adults with generalized anxiety disorder. *Psychological Assessment, 15*, 184-192.

Berger, A., Small B., Forsell, Y., & Winblad, B. (1998). Preclinical symptoms of depression in very old age: A prospective longitudinal study. *The American Journal of Psychiatry, 155*, 1039.1043.

Blazer, D. G., Burchett, B., Service, C., & George, L. K. (1991). The association of age and depression among older adults: An epidemiological exploration. *Journal of Gerontology, 46*, 210-215.

Boyd, D. & Bee, H. (2006). *Lifespan development* (4th ed.). Belmont, CA: Wadsworth.

Brodaty, H., Luscombe, G., & Parker, G. (1997). Increased rate of psychosis and psychomotor change in depression with age. *Psychological Medicine, 27*, 1205-1213.

Brodaty, H., Thompson, C., & Mitchell, P. (2005). Age and gender in the phenomenology of depression. *American Journal of Geriatric Psychiatry, 13*, 589-596.

Brodaty, H., Peters, K., & Boyce, P. (1991). Age and depression. *Journal of Affective Disorders, 23*, 137-149.

Brody, D. W., & Serby, M. (2004). Treating adult depression: A guide to a successful strategy. *Cortlandt Forum, 17*, 28-41. Retrieved from Academic Search Complete database August 30, 2008.

Brink, T. A., Yesavage, J. A., Lum, O., Heersema, P., Adey, M., & Rose, T. L. (1982). Screening tests for geriatric depression. *Clinical Gerontologist*, *1*, 37-44.

Brownhill, S., Wilhelm, K., Barclay, L., & Parker, G. (2002). Detecting depression in men: A matter of guesswork. *International Journal of Men's Health, 1*, 259-280.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., & Kaemmer, B. (1992). *Manual for administration, scoring, and interpretation of the Minnesota Multiphasic Personality Inventory for Adolescents.* Minneapolis, MN: University of Minnesota Press.

Cahill, J, Barkham, M., Stiles, W. B., & Twigg, E. (2006). Convergent validity of the CORE measures with measures of depression for clients in cognitive therapy for depression. *Journal of Counseling Psychology, 53*, 253-259.

Callahan, C. M., Hui, S. L., Nienaber, N. A., & Musick, B. S. (1994). Longitudinal study of depression and health service use among older adults primary care patients. *Journal of the American Geriatrics Society, 42*, 833-888.

Cassidy, E. L., Lauderdale, S., & Sheikh, J, I. (2005). Mixed anxiety and depression in older adults: Clinical characteristics and management. *Journal of Geriatric Psychology and Neurology, 18*, 83-88.

Cavanaugh, J. C., & Blanchard-Fields, F. (2002). *Adult development and aging* (4th ed.). Belmont, CA: Wadsworth.

Coolidge, F. L., & Merwin, M. M. (1992). Reliability and validity of the Coolidge Axis-II Inventory: A new inventory for the assessment of personality disorders. *Journal of Personality Assessment, 59,* 223-238.

Davidson, T. E., McCabe, M. P, & Mellor, D. (2009). An examination of the "gold standard" diagnosis of major depression in aged-care settings. *American Journal of Geriatric Psychiatry, 17*(5), 359-367.

Dennis, R. E., Boddington, S. J. A., Funnell, N. J. (2007). Self-report measures of anxiety: Are they suitable for older adults? *Aging & Mental Health, 11*, 668-677.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffen, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49,* 71-75.

Dixon, A. L. (2007). Mattering in the later years: Older adults' experiences of mattering to others, purpose in life, depression, and wellness. *Adultspan Journal, 6*, 83-95.

Djernes, J. K. (2006). Prevalence and predictors of depression in populations of elderly: A review. *Acta Psychiatrica Scandinavica, 113*, 372-387.

Doris, A., Ebmeier, K., & Shajahan, P. (1999). Depressive illness. *The Lancet, 354*, 1369-1375. Retrieved from Academic Search Complete database August 30, 2008.

Dowd, E. T. (1998). Review of the Beck Anxiety Inventory. In J. C. Impara, & B. S. Plake (Eds.), *The thirteenth mental measurements yearbook.* Retrieved September 8, 2008 from the Mental Measurement Yearbook database.

Dunn, V., & Sacco, W. (1989). Psychometric evaluation of the Geriatric Depression Scale and the Zung Self-rating Depression Scale using an elderly community sample. *Psychology and Aging, 4*, 125-126.

First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W. (n.d.). Structured Clinical Interview for DSM-IV (SCID). Welcome to the SCID webpage!!! Retrieved August 31, 2008 from http://cpmcnet.columbia.edu/dept/scid/.

First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W. (2002). *User's guide for the Structured Clinical Interview for DSM-IV-TR Axis-I Disorders - Research Version - (SCID-I for DSM-IV-TR).* New York: Biometrics Research, New York State Psychiatric Institute.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2007). *Structured Clinical Interview for DSM-IV-TR Axis-I Disorders, Research Version, Non-patient Edition (SCID-I/NP).* New York: Biometrics Research, New York State Psychiatric Institute.

Folstein, M. F., Anthony, J. C., Parhad, I., Duffy, B., & Gruenberg, E. M. (1985). The meaning of cognitive impairment in the elderly. *Journal of the American Geriatric Society, 33,* 228-235.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189-198.

Folstein, M. F., Folstein, S. E., & Fanjiang, G. (2001). *Mini-Mental State Examination clinical guide.* Lutz, FL: Psychological Assessment Resources, Inc.

Frisch, M. B. (1994). *Manual and treatment guide for the Quality of Life Inventory.* Minneapolis, MN: National Computer Systems, Inc.

Fuentes, K., & Fox, B. (2000). Assessment of anxiety in older adults: a community-based survey and comparison with younger adults. *Behavior Research and Therapy, 38*, 297-309.

Gallagher, D. (1986). The Beck Depression Inventory and older adults: Review of its development and utility. *Clinical Gerontologist, 5*, 149-163.

Gatz, M. (2000). Variations on depression in later life. In S. H. Qualls & N. Abeles (Eds.), *Psychology and the aging revolution: How we adapt to longer life* (pp. 239-254). Washington, DC, US: American Psychological Association.

Gatz, M., & Fiske, A., (2003). Aging women and depression. *Professional Psychology: Research and Practice*, *34*, 3-9.

Gregory, R. J. (2004). Validity and test development. *Psychological testing: History, principles, and applications.* (Chapter 4, pp. 96-137). Boston, MA: Pearson Education Group.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry, 23,* 56-62.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7,* 3, 238-247.

Holmes, W. C. (1998). A short, psychiatric, case-finding measure for HIV seropositive outpatients: Performance characteristics of the 5-item Mental Health Subscale of the SF-20 in a male, seropositive sample. *Medical Care, 36,* 237-243.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V. S. (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56,* 885-892.

Hsiao, J. K., Bartko, J. J., Potter, W. Z. (1989). Receiver operating characteristic methods and psychiatry. *Archives of General Psychiatry, 46,* 664-667.

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

Hubley, A. M. (1998). Hubley Depression Scale for Older Adults (HDS-OA). Unpublished manuscript. University of Northern British Columbia, Prince George, BC.

Hubley, A. M., Mangaoang, M., Burke, S., Ho, C., Ang, P., Myers, S., & Chiu, D. (August, 2009). Validation of a new screen for depression in older adults. Poster presented at the *American Psychological Association 117th Annual Convention*, Toronto, ON.

Hubley, A., M. & Zumbo, B. D. (1996). A dialectic on validity: Where we have been and were we are going. *The Journal of General Psychology, 123*, 207-215.

Huffstetler, B. (2001). Depression in older adults: Pervasive or preventable? *Adultspan Journal, 3*, 61-70.

Joseph, S., Lewis, C. A., & Olsen, C. (1996). Convergent validity of the Depression-Happiness Scale with measures of depression. *Journal of Clinical Psychology, 52*, 551-554.

Kabacoff, R. I., Segal, D. L., Hersen, M., & Van Hasselt, V. B. (1997). Psychometric properties and diagnostic utility of the Beck Anxiety Inventory and the State-Trait Anxiety Inventory with older adult psychiatric outpatients. *Journal of Anxiety Disorders 11*, 33–47.

Katona, C., & Livingston, G. (2000). Impact of screening old people with physical illness for depression. *The Lancet, 356*, 91-92.

Katz, I. R. (2000). Opportunities for redefining late life depression ab initio. In R. L. Rubenstein, M. Moss, & M. H. Kleban (Eds.), *The many dimensions of aging* (pp. 164-172). New York: Springer Publishing Company.

Keppel, G., & Wickens, T. D. (2004). *Design & analysis: A researcher's handbook.* Upper Saddle River, N. J.: Pearson Prentice Hall.

Kraaij, V. & Garnefski, N. (2002). Negative life events and depressive symptoms in late life: Buffering effects of parental and partner bonding? *Personal Relationships, 9*, 205-214.

Kupfer, D. J., Kuhl, E. A., & Regier, D. A. (2009). Research for improving diagnostic systems: Consideration of factors related to later life development. *American Journal of Geriatric Psychiatry, 17*, 355-358.

Lewinsohn, P. M., Seeley, J. R., Roberts, R. E., & Allen, N. B. (1997). Center for Epidemiologic Studies Depression scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychology and Aging, 12*, 277-287.

Lockwood, K., Alexopoulos, G., & van Gorp, W. (2002). Executive function in geriatric depression. *American Journal of Psychiatry, 159*, 1119-1123.

Low, G. D., & Hubley, A. M. (2007). Screening for depression after cardiac events using the Beck Depression Inventory-II and the Geriatric Depression Scale. *Social Indicators Research, 82*, 527-543.

Lyness, J. M., Cox, C., & Curry, J. (1995). Older age and the underreporting of depressive symptoms. *Journal of the American Geriatric Society, 43*, 216-221.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*, 362-365.

Marquez, D. X., McAuley, E, Motl, R. W., & Elavsky, S. (2006). Validation of Geriatric Depression Scale-5 scores among sedentary older adults. *Educational and Psychological Measurement, 66*, 667-675.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Berkovee, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behavioral Research and Therapy, 28,* 487-495.

Myers, J. E. (1989). *Adult children and aging parents.* Alexandra, VA: American Association for Counseling and Development.

National Institute of Mental Health. (2003). *Older adults: Depression and suicide facts.* Retrieved August 16, 2008 from http://www.nimh.nih.gov/publicat/elderlydepsuicide.cfm.

National Institute of Mental Health. (2007). *Depression.* Retrieved August 18, 2008 from http://www.nimh.nih.gov/health/publications/depression/complete-publication.shtml.

Norris, M. P., Snow-Turek, A. L., & Blankenship, L. (1995). Somatic depressive symptoms in the elderly: Contribution or confound? *Journal of Clinical Geropsychology, 1*, 5-17.

Olin, J. T., Schneider, L. S., Eaton, E. M., & Zemansky, M. F. (1992). The Geriatric Depression Scale and the Beck Depression Inventory as screening instruments in an older adult outpatient population. *Psychological Assessment, 4*, 190-192.

Oliver, S., & Toner, B. (1990). The influence of gender role typing on the expression of depressive symptoms. *Sex Roles, 22*, 775-790.

Onega, L. L. (2006). Assessment of psychoemotional and behavioral status in patients with dementia. *Nursing Clinics of North America, 41*, 23-41.

Onega, L. L. (2008). Content validation of the Depressive Symptom Assessment for Older Adults. *Issues in Mental Health Nursing, 29,* 873-894.

Osman, A., Bagge, C. L., Gutierrez, P. M., Konick, L. C., Kopper, B. A., & Barrios, F. X. (2001). The Suicidal Behaviors Questionnaire – Revised (SBQ-R): Validation with clinical and non-clinical samples. *Assessment, 8,* 443-454.

Osman, A., Kopper, A., Barrios, F., & Gutierrez, P. M. (2004). Reliability and validity of the Beck Depression Inventory-II with adolescent psychiatric inpatients. *Psychological Assessment, 16*, 120-132.

Pitt, B. (1998). Loss in late life. *British Medical Journal, 316*, 1452-1454.

Radoff, L. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401.

Rapp, S. R., Parisi, S. A., Walsh, D. A., & Wallace, C. E. (1988). Detecting depression in elderly medical in-patients. *Journal of Consulting and Clinical Psychology, 56*, 509-513.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57,* 1069-1081.

Sable, J., Dunn, L., & Zisook, S. (2002). Late-life depression: How to identify its symptoms and provide effective treatment. *Geriatrics, 57*, 18-35.

Segal, D. L, Coolidge, F. L., Cahill, B. S., & O'Riley, A. A. (2008). Psychometric properties of the Beck Depression Inventory-II (BDI-II) among community-dwelling older adults. *Behavior Modification, 32*, 3-20.

Serby, M., & Yu, M. (2003). Overview: Depression in the elderly. *Mount Sinai Journal of Medicine, 70*, 38-44.

Shear, M. K., Greeno, C., Kang, J., & Ludewig. D. (2000). Diagnosis of nonpsychotic patients in community clinics. *American Journal of Psychiatry, 157*, 581-587.

Sheikh, J. I, & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Geronotologist, 5*, 165-173.

Shultz, K. S., Riggs, M. L., & Kottke, J. L. (1998). The need for an evolving concept of validity in industrial and personnel psychology: Psychometric, legal, and emerging issues. *Current Psychology,* 17, 265-77.

Shultz, K. S., & Whitney, D. J. (2005). Exploratory and confirmatory factor analysis. *Measurement theory in action: Case studies and exercises.* (Module 18, pp. 313-326). Thousand Oaks, CA: Sage Publications.

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

Snyder, A. G., Stanley, M. A., Novy, D. M., & Averill, P. M. (2000). Measures of depression in older adults with generalized anxiety disorder: A psychometric evaluation. *Depression and Anxiety, 11*, 114-120.

Spielberger, C. D. (1983). *State-trait anxiety inventory (form Y)*. Redwood City, CA: Mind Garden.

Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1999). Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *Journal of Clinical Psychology, 55*, 117-128.

Steer, R. A., Rissmiller, D. J., & Beck, A. T. (2000). Use of the Beck Depression Inventory-II with depressed geriatric inpatients. *Behavior Research and Therapy, 38*, 311-318.

Steiner, J. L., Tebes, J. K., & Sledge, W. H. (1995). A comparison of the structured clinical interview for DSM-III-R and clinical diagnoses. *Journal of Nervous and Mental Disease, 183*, 365-369.

Stickle, F., & Onedera, J. D. (2006). Depression in older adults. *Adultspan Journal, 5*, 36-46.

Strupp, H. H., Horowitz, L. M., & Lambert, M. J. (Eds.) (1997). *Measuring patient changes in mood, anxiety, and personality disorders: Toward a core battery.* Washington, DC: American Psychological Association.

Swets, J. A., Dawes, R. M., Monahav, J. (2000). Better decisions through science. *Scientific American, 283,* 82-87.

Teper, E. & Thomas, A. (2006). Is depression different in older adults? *Aging Health, 2,* 905-915.

Teri, L., & Wagner, A. W. (1991). Assessment of depression in patients with Alzheimer's disease: Concordance among informants. *Psychology and Aging, 6*, 280-285.

Uncapher, H., Gallagher-Thomas, D., Osgood, N. & Bongar, B. (1998). Hopelessness and suicidal ideation in older adults. *The Gerontologist, 38*, 62-70.

van Marwijk, H. W. J., Wallace, P., De Bock, G. H., Hernans, J. O., Kptein, A. A., & Mulder, J. D. (1995). Evaluation of the feasibility, reliability and diagnostic value of shortened versions of the Geriatric Depression Scale. *British Journal of General Practice, 45*, 195-199.

Vink, D., Aartsen, M. J., & Schoevers, R. A. (2008). Risk factors for anxiety and depression in the elderly: A review. *Journal of Affective Disorders*, *106*, 29-44.

Wancata, J., Alexandrowicz, R., Marquart, B., Weiss, M., & Friedrich, F. (2006). The criterion validity of the Geriatric Depression Scale: A systematic review. *Acta Psychiatrica Scandinavica, 114*, 398-410.

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

Wassertheil-Smoller, S. (1990). *Biostatistics and epidemiology: A primer for health professional*s. Springer-Verlag: NY.

Watson D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063-1070.

Waza, K., Graham, A. V., Zyzanski, S. J., & Inoue, K. (1999). Comparison of symptoms in Japanese and American depressed primary care patients. *Family Practice, 16*, 528-533.

Wisocki, P. A., Handen, B., & Morse, C. (1986). The Worry Scale as a measure of anxiety among homebound and community active elderly. *Behavioral Therapist, 5*, 91-95.

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1983). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research, 17*, 36-40.

Zung, W. W. K. (1965). A self-rating depression scale. *Archives of General Psychiatry, 12*, 63-60.

Appendix A – Letter of Initial Contact and Consent to be Contacted for Depressed Group

THE UNIVERSITY OF BRITISH COLUMBIA

**Department of Educational and Counselling
Psychology, and Special Education**
2125 Main Mall
Vancouver, BC, Canada, V6T 1Z4
Tel: (604) 822-9223
Fax: (604) 822-3302

INFORMATION LETTER:

Depressed Adults Ages 60+
Investigating Criterion Convergent, and Discriminant Validity of the Hubley Depression Scale for Older Adults (HDS-OA)

Purpose of the study. <u>Men and women ages 60 years and older</u> are asked to participate in a research project to evaluate a new questionnaire developed to screen for depression in older adults. This study is being conducted by Sherrie Myers, a Counselling Psychology graduate student at The University of British Columbia and Dr. Anita Hubley, a Professor at The University of British Columbia. This study will help determine whether this new shorter questionnaire is a helpful measure of the symptoms of depression in older adults.

Participation in the study. We are currently seeking <u>individuals who have a current diagnosis of depression</u> to volunteer for this research study. Participants will be asked to complete a brief interview about thinking abilities and memory, two short questionnaires that ask about current mood, and one short questionnaire that asks about current responses to stressful situations. This will take about <u>30 minutes</u> in total. The study session can take place in your home, at the hospital, or at the University of British Columbia. Participants will receive a $10 gift card as a token of appreciate for involvement in this research.

If you have any questions or would like to take part in this study:

1) Please contact <u>Sherrie Myers at 604-822-xxxx</u> (Adult Development and Psychometrics Laboratory where you can leave a message) or ubc_adultdevelopment@hotmail.com, OR

2) Provide your contact phone number below information and you will be contacted within one week to discuss participation in the study.

Thank you!

CONSENT TO BE CONTACTED:

I have read the above information and I freely consent to be contacted by phone to learn more about the study entitled "Criterion Convergent, and Discriminant Validation of the Hubley Depression Scale for Older Adults (HDS-OA)" and to discuss my potential participation. I acknowledge receipt of a copy of this information letter.

_____   _____   _____

**Signature of   Name (please print)   Date
Potential Participant**

Contact Phone Number: _____

Study Title: Criterion, Convergent, and Discriminant Validation of the Hubley Depression Scale for Older Adults (HDS-OA)

Principal Investigator: Dr. Anita Hubley, Professor in the Dept. of Educational and Counselling Psychology, and Special Education, UBC. Co-investigator: Sherrie Myers, Dept. of Educational and Counselling Psychology, and Special Education, UBC.

Adapted from A. Hubley (personal communication, June, 2008)

Appendix B – Recruitment Advertisement for Non-depressed Group

UBC THE UNIVERSITY OF BRITISH COLUMBIA

Department of Educational and Counselling
Psychology, and Special Education
2125 Main Mall
Vancouver, BC, Canada, V6T 1Z4
Tel: (604) 822-9223
Fax: (604) 822-3302

SEEKING NON-DEPRESSED ADULTS - AGES 60+

<u>Men and women ages 60 years and older</u> are asked to participate in an important research project to evaluate a new questionnaire developed to screen for depression in older adults. This study is being conducted by Sherrie Myers, a Counselling Psychology graduate student at The University of British Columbia and Dr. Anita Hubley, a Professor at The University of British Columbia. This study will help determine whether this new shorter questionnaire is a helpful measure of symptoms of depression in older adults.

We are currently <u>seeking non-depressed individuals from the community</u> that we can compare to a sample of individuals who are diagnosed with depression.

Participants in this study will be asked to complete a brief interview about current cognitive functioning and memory, two short questionnaires that ask about current mood, one short questionnaire that asks about current responses to stressful situations, and participate in a longer interview about mood and medical history. This will take about <u>75 minutes</u> in total. The study session can take place in your home or The University of British Columbia. Participants will receive a $20 gift card as a token of appreciate for their involvement in this study.

If you have any questions or would like to discuss eligibility for the study, please contact <u>Sherrie Myers</u> at <u>604-822-xxxx</u> (Adult Development and Psychometrics Laboratory where you can leave a message) or ubc_adultdevelopment@hotmail.com.

NON-DEPRESSED PARTICIPANTS NEEDED FOR STUDY

Thank you!

Study Title: Criterion, Convergent, and Discriminant Validation of the Hubley Depression Scale for Older Adults (HDS-OA). Principal Investigator: Dr. Anita Hubley, Professor in the Dept. of Educational and Counselling Psychology, and Special Education, UBC. Co-investigator: Sherrie Myers, Dept. of Educational and Counselling Psychology, and Special Education, UBC.

Adapted from A. Hubley (personal communication, June, 2008)

Appendix C – Demographics Information Sheet

 THE UNIVERSITY OF BRITISH COLUMBIA

**Department of Educational and Counselling
Psychology, and Special Education**
2125 Main Mall
Vancouver, BC, Canada, V6T 1Z4
Tel: (604) 822-9223

**Criterion, Convergent, and Discriminant Validation of the Hubley Depression Scale
for Older Adults (HDS-OA)**

Participant Demographics Information Sheet

ID Number (researcher will enter): _____

Age: _____   Date of Birth (dd/mm/yy): _____

Gender (please circle): Female / Male

What is your *primary* ethnic/racial/cultural background? (please check one box only):

☐ Aboriginal/First Nations

☐ Black (e.g., African American/Canadian)

☐ East Asian (e.g., Chinese, Japanese, Korean)

☐ Hispanic (e.g., Latino, Mexican)

☐ South Asian / Middle Eastern (e.g., Indian, Arabic)

☐ White (e.g., Caucasian, Anglo)

☐ Other: (please specify: _____)

(note: Please do <u>not</u> record *nationality*, such as "Canadian", for this question.)

Marital Status (place a check in the appropriate box):

☐ Married/Common-law

☐ Separated/Divorced

☐ Single – never married

☐ Widowed

Education (please check your highest level of education):

☐ Elementary school – not completed

☐ Elementary school – completed

☐ High school – not completed

☐ High School – completed

☐ Some trade, technical, business or community college – not completed

☐ Diploma or certificate from trade, technical, business, or community college –
completed

☐ Some university – not completed

☐ Bachelor Degree – completed

☐ Master's Degree – completed

☐ Doctoral Degree – completed

In general, would you say that your health is (place a check in the appropriate box):

☐ Excellent

☐ Very Good

☐ Good

☐ Fair

☐ Poor

Current anti-depressant medication usage? (please circle): Yes/No

Appendix D – Informed Consent Form for Depressed Group

THE UNIVERSITY OF BRITISH COLUMBIA

Department of Educational and Counselling Psychology, and Special Education
2125 Main Mall
Vancouver, BC, Canada, V6T 1Z4
Tel: (604) 822-9223
Fax: (604) 822-3302

<u>Subject Information and Consent Form – Depressed Group Participants</u>

Criterion, Convergent, and Discriminant Validation of the Hubley Depression Scale for Older Adults (HDS-OA)

Principal Investigator. Dr. Anita Hubley, Professor in the Department of Educational and Counselling Psychology, and Special Education at The University of British Columbia is the faculty supervisor overseeing this research. Telephone: (604) 822-xxxx.

Co-Investigator. Sherrie Myers, a graduate student in the Department of Educational and Counselling Psychology, and Special Education at The University of British Columbia, is the co-investigator and is conducting this research for her Master's thesis in Counselling Psychology. Telephone: (604) 822-xxxx (Adult Development and Psychometrics Laboratory).

1. Introduction. This study is investigating a new screening measure for depression in older adults, and you are being invited to take part in this research study because you may be an older adult who is currently diagnosed with depression.

2. Voluntary Participation. Your participation is entirely voluntary. You have the right to refuse to participate in this study. Before you decide, it is important for you to understand what might be involved. This informed consent will confirm that you understand the purpose of this study, your role in this study and what you are being asked to do, and your rights as a participant. If you decide to participate, your decision is not binding and you may choose to withdraw from the study at any time without any negative consequences to the medical care, education, or other services you may receive from this clinic or this hospital.

3. Purpose. The purpose of this study is to compare a new, brief questionnaire measuring depression called the Hubley Depression Scale for Older Adults (HDS-OA) with an older, widely used questionnaire, called the Geriatric Depression Scale (GDS). This will allow us to determine whether the shorter HDS-OA is a helpful measure of the symptoms of depression in older adults. To do this, we are conducting this study with both a non-depressed community sample and a clinically diagnosed depressed sample of older adults.

4. What is Required. You are being asked to complete a brief questionnaire about your current thinking abilities and memory, two short questionnaires that ask about your current mood, and one short questionnaire about current responses to stressful situations.

5. Release of Research Specific Medical Information. Your consent for the release of medical information from your treating physician or psychologist is requested during the time of the study only to confirm a diagnosis of depression and the absence of a psychotic disorder, and to confirm that the depression is not related to a substance induced or general medical condition. Only the questionnaire results from those who meet these criteria will be included in the study.

6. Length of Participation. Completion of the questionnaires should take approximately 30 minutes during one meeting with the co-investigator. However, the questions are not timed and you may take as long as you want on each one. You may refuse to answer any questions.

7. Benefits. Although no direct benefit is anticipated from study participation, you may derive satisfaction in knowing that you have participated in a study that may help produce a better screen for depression in older adults.

8. Risks. This study focuses on experiences of negative mood. In thinking about these experiences while answering the questionnaires, you may experience some distress during or after the session. It is suggested that you speak with their primary/treating physician, psychologist, or psychiatrist if emotional distress arises.

9. Rights and Compensation. By signing this form, you do not give up any of your legal rights and you do not release the researchers from their legal and professional duties. There will be no costs to you for participation in this study and you will not be charged for any research procedures. We can offer reimbursement for parking costs incurred during study participation and you will receive a $10 gift certificate as a token of appreciation for your involvement in the study.

10. Confidentiality. Your confidentiality will be respected. Information that discloses your identity will not be released without your consent unless required by law or regulation. However, research records and medical records identifying you may be inspected in the presence of the investigator or his or her designate, by representatives of the UBC-PHC Research Ethics Board for the purposes of monitoring the research. No records that identify you by name or initials will be allowed to leave the investigator's office. To protect confidentiality, code numbers will be assigned to your questionnaires, and all results will be kept in a locked room at UBC. The master list with participant's names and code numbers will only be available to Dr. Anita Hubley, Sherrie Myers, and the UBC Adult Development and Psychometrics Lab Manager, Lara Russell.

11. Limit to Confidentiality. One limit to confidentiality is that the researchers will be ethically obligated to report if it is revealed that you may be at imminent risk of harm to yourself or others, and to solicit help for you from a mental health professional.

12. Subject's Right to Withdraw. You have the right to forego answering any specific question. You also have the right to withdraw your consent and withdraw from the study, without providing a reason and without any negative consequences.

13. Rights As a Research Subject (ICH-GCP 4.8.10 (q)). If you have any concerns about your rights as a research subject and/or your experiences while participating in this study, contact the 'Research Subject Information Line in the University of British Columbia Office of Research Services' at 604-822-8598" or the Chair of the UBC-PHC Research Ethics Board at 604-682-2344 ext 63496.

14. Important Information. You will be provided a copy of this signed and dated informed consent sheet for your files. If required in future, it is the researcher's responsibility to produce the signed copy of this form demonstrating that you consented to be a participant in this research.

15. Contact Persons. Any questions you may have about this study may be directed to Sherrie Myers at (604) 822-xxxx (Adult Development and Psychometrics Laboratory where you can leave a message) or Dr. Anita Hubley at (604) 822-xxxx.

16. Feedback. If you would like a summary of the overall study results to be sent to you when the research is completed, please check here ☐.

Full mailing address or email address: _____
(if checked above)   _____

                                          _____

Consent.
☐ I have read the above description of the study "Criterion Convergent, and Discriminant Validation of the Hubley Depression Scale for Older Adults (HDS-OA)".
☐ I have been given an explanation of the purpose and scope of the research and the conditions of my participation.
☐ My signature confirms that I agree to participate in the study and that I have been offered a copy of this consent form.

_____
Printed name of subject   Signature   Date

_____
Printed name of person   Signature   Date
obtaining consent

_____
Treating Physician or Psychologist's Name

Adapted from A. Hubley (personal communication, May, 2008)

Appendix E – Depression Resource List

Referral List of Community Resource Services

Criterion, Convergent, and Discriminant Validation of the Hubley Depression Scale for Older Adults (HDS-OA): Community Subjects

If you find you are experiencing negative mood, sadness, or depression and you would like to know more about these issues or if you would like to speak with someone to discuss concerning emotional issues, the following is a list of resources that are available in the greater Vancouver area as well as on-line:

1) Mood Disorders Association of B.C.   604-873-0103
    202-2250 Commercial Drive, Vancouver   www.mdabc.net

2) BC Partners for Mental Health & Addictions Information
                                             http://www.heretohelp.bc.ca/about

3) National Institute of Mental Health   http://www.nimh.nih.gov/

4) Changeways Clinic   604-871-0490
    2525 Willow St., Suite 509, Vancouver   www.changeways.com

5) Family Services of Greater Vancouver   604-731-4951
    1616 West 7th, Vancouver   (Sliding fee scale)

6) Jewish Family Service Agency   604-257-5151
     950 West 41st, Vancouver   (Sliding fee scale)

7) B.C. Association of   1-800-909-6303
    Clinical Counsellors (Referrals/Information) www.bc-counsellors.org

8) British Columbia Psychological Association 604-730-0522
    (Referral Service)   www.psychologists.bc.ca/referral.html

For immediate assistance:

1) Crisis Centre: 24 hour distress line
    www.crisiscentre.bc.ca   604-872-3311 or 1-800-784-2433

2) Mental Health Emergency   24 hours
    Services   604-874-7307

A. Hubley (personal communication, May, 2008)

Appendix F – Informed Consent Form for Non-depressed Group

![UBC logo] **THE UNIVERSITY OF BRITISH COLUMBIA**

**Department of Educational and Counselling
Psychology, and Special Education**
2125 Main Mall
Vancouver, BC, Canada, V6T 1Z4
Tel: (604) 822-9223
Fax: (604) 822-3302

**Subject Information and Consent
Form – Non-Depressed Group Participants**

**Criterion, Convergent, and Discriminant Validation of the Hubley Depression Scale
for Older Adults (HDS-OA)**

Principal Investigator. Dr. Anita Hubley, Professor in the Department of Educational
and Counselling Psychology, and Special Education at The University of British
Columbia is the faculty supervisor overseeing this research. Telephone: (604) 822-xxxx.

Co-Investigator. Sherrie Myers, a graduate student in the Department of Educational and
Counselling Psychology, and Special Education at The University of British Columbia, is
the co-investigator and is conducting this research for her Master's thesis in Counselling
Psychology. Telephone: (604) 822-xxxx (Adult Development and Psychometrics
Laboratory).

1. Introduction. This study is investigating a new screening measure for depression in
older adults, and you are being invited to take part in this research study because you may
be a non-depressed older adult.

2. Voluntary Participation. Your participation is entirely voluntary. You have the right
to refuse to participate in this study. Before you decide, it is important for you to
understand what might be involved. This informed consent will confirm that you
understand the purpose of this study, your role in this study and what you are being asked
to do, and your rights as a participant. If you decide to participate, your decision is not
binding and you may choose to withdraw from the study at any time without any negative
consequences.

3. Purpose. The purpose of this study is to compare a new, brief questionnaire measuring
depression called the Hubley Depression Scale for Older Adults (HDS-OA) with an
older, widely used questionnaire, called the Geriatric Depression Scale (GDS). This will
allow us to determine whether the shorter HDS-OA is a helpful measure of the symptoms
of depression in older adults. To do this, we are conducting this study with both a non-
depressed community sample and a clinically diagnosed depressed sample of older
adults.

4. What is Required, You are being asked to complete a brief questionnaire about your current thinking abilities and memory, two short questionnaires that ask about your current mood, one short questionnaire about current responses to stressful situations, as well as participate in an interview about your mood and medical history.

5. Length of Participation Completion of the questionnaires and interview should take approximately 75 minutes in total. The questions are not timed, however, and you may take as long as you want on any item. You may refuse to answer any questions.

6. Benefits. Although no direct benefit is anticipated from study participation, you may derive satisfaction in knowing that you have participated in a study that may help produce a better screen for depression in older adults.

7. Risks. This study focuses on experiences of negative mood. In thinking about these experiences while answering the questionnaires and speaking about them during the interview, you may experience some distress during or after the session. As a result, you will be provided with information about community counselling services and resources. At the end of the interview, if you have concerns or significant symptoms of depression, it will be suggested that you speak with your physician about your concerns and experience of negative mood. This information will not otherwise be reported to family or medical personnel unless it is determined that there is a risk of imminent harm to yourself or others (see section on Limits to Confidentiality below).

8. Rights and Compensation. By signing this form, you do not give up any of your legal rights and you do not release the researchers from their legal and professional duties. There will be no costs to you for participation in this study and you will not be charged for any research procedures. We can offer reimbursement for parking costs incurred during study participation and you will receive a $20 gift certificate as a token of appreciation for your involvement in the study.

9. Confidentiality. Your confidentiality will be respected. Information that discloses your identity will not be released without your consent unless required by law or regulation. However, research records and medical records identifying you may be inspected in the presence of the investigator or his or her designate, by representatives of the UBC-PHC Research Ethics Board for the purposes of monitoring the research. No records that identify you by name or initials will be allowed to leave the investigator's office.

To protect confidentiality, code numbers will be assigned to your questionnaires, and all results will be kept in a locked room at UBC. The master list with participant's names and code numbers will only be available to Dr. Anita Hubley, Sherrie Myers, and the UBC Adult Development and Psychometrics Lab Manager, Lara Russell.

10. Limit to Confidentiality. One limit to confidentiality is that the researchers will be ethically obligated to report if it is revealed that you may be at imminent risk of harm to yourself or others, and to solicit help for you from a mental health professional.

11. Subject's Right to Withdraw. You have the right to forego answering any specific question. You also have the right to withdraw your consent and withdraw from the study, without providing a reason and without any negative consequences.

12. Rights As a Research Subject (ICH-GCP 4.8.10 (q)). If you have any concerns about your rights as a research subject and/or your experiences while participating in this study, contact the 'Research Subject Information Line in the University of British Columbia Office of Research Services' at 604-822-8598" or the Chair of the UBC-PHC Research Ethics Board at 604-682-2344 ext 63496.

13. Important Information. You will be provided a copy of this signed and dated informed consent sheet for your files. If required in future, it is the researcher's responsibility to produce the signed copy of this form demonstrating that you consented to be a participant in this research.

14. Contact Persons. Any questions you may have about this study may be directed to Sherrie Myers at (604) 822-xxxx (Adult Development and Psychometrics Laboratory where you can leave a message) or Dr. Anita Hubley at (604) 822-xxxx.

15. Feedback. If you would like a summary of the overall study results to be sent to you when the research is completed, please check here ☐.

Full mailing address or email address: _____
(if checked above)   _____
                                        _____

Consent.
☐ I have read the above description of the study "Criterion Convergent, and Discriminant Validation of the Hubley Depression Scale for Older Adults (HDS-OA)".
☐ I have been given an explanation of the purpose and scope of the research and the conditions of my participation.
☐ My signature confirms that I agree to participate in the study and that I have been offered a copy of this consent form.

_____
Printed name of subject   Signature   Date

_____
Printed name of person   Signature   Date
obtaining consent

Adapted from A. Hubley (personal communication, May, 2008)

Appendix G – Certificates of Ethical Approval to Conduct Research

UBC · PROVIDENCE HEALTH CARE Research Institute

UBC-Providence Health Care Research Institute
Office of Research Services
11th Floor Hornby Site - SPH
c/o 1081 Burrard St.
Vancouver, BC V6Z 1Y6
Tel: (604) 806-8567
Fax: (604) 806-8568

ETHICS CERTIFICATE OF EXPEDITED APPROVAL

PRINCIPAL INVESTIGATOR:	DEPARTMENT:	UBC-PHC REB NUMBER:
Anita M. Hubley	UBC/Education/Educational & Counselling Psychology, and Special Education	H09-00304

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
Vancouver Coastal Health (VCHRI/VCHA)	Vancouver General Hospital
Providence Health Care	Mount Saint Joseph Hospital
Providence Health Care	St. Paul's Hospital
Vancouver Coastal Health (VCHRI/VCHA)	Vancouver Community

Other locations where the research will be conducted:
Subject's home University of British Columbia

COINVESTIGATOR(S):
Sherrie L. Myers
Martha L. Donnelly

SPONSORING AGENCIES:
British Columbia Network for Aging Research - Michael Smith Foundation for Health Research (MSFHR)
UBC Faculty of Education

PROJECT TITLE:
Criterion, Convergent, and Discriminant Validation of the Hubley Depression Scale for Older Adults (HDS-OA)

THE CURRENT UBC-PHC REB APPROVAL FOR THIS STUDY EXPIRES: April 13, 2010

The UBC-PHC Research Ethics Board Chair or Associate Chair, has reviewed the above described research project, including associated documentation noted below, and finds the research project acceptable on ethical grounds for research involving human subjects and hereby grants approval.

DOCUMENTS INCLUDED IN THIS APPROVAL:		APPROVAL DATE: April 13, 2009
Document Name	**Version**	**Date**
Protocol:		
Study Protocol - HDS-OA Validation	1	February 21, 2009
Consent Forms:		
Consent Form - Non-Depressed Group	2	March 28, 2009
Consent Form - Depressed Group	2	March 28, 2009
Advertisements:		
Recruitment Ad - Non-Depressed Group	1	February 21, 2009
Recruitment Ad - Depressed Group	1	February 21, 2009
Questionnaire, Questionnaire Cover Letter, Tests:		
Demographics Information Sheet	1	February 21, 2009
Geriatric Depression Scale (GDS)	1	January 1, 1983
Beck Anxiety Inventory (BAI)	1	January 1, 1990
Hubley Depression Scale for Older Adults (HDS-OA)	1	January 1, 1998

Page 1 of 2

Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID-I/NP) Interview Overview	1	January 1, 2002
Mini-Mental State Examination (MMSE)	1	May 1, 2008
Letter of Initial Contact:		
Letter of Initial Contact - Depressed Group	2	March 28, 2009
Other Documents Acknowledged		
Referral List of Community Resource Services	1	February 21, 2009
UBC Research Committee Study Approval Letter	1	February 12, 2009

CERTIFICATION:

1. The membership of the UBC-PHC REB complies with the membership requirements for research ethics boards defined in Part C Division 5 of the Food and Drug Regulations of Canada.
2. The UBC-PHC REB carries out its functions in a manner fully consistent with Good Clinical Practices.
3. The UBC-PHC REB has reviewed and approved the research project named on this Certificate of Approval including any associated consent form and taken the action noted above. This research project is to be conducted by the principal investigator named above at the specified research site(s). This review of the UBC-PHC REB have been documented in writing.

Approval of the UBC-PHC Research Ethics Board or Associate Chair, verified by the signature of one of the following:

Dr. Kuo-Hsing Kuo, Chair	Dr. J. Kernahan, Associate Chair	Dr. I. Fedoroff, Associate Chair

ETHICS CERTIFICATE OF EXPEDITED APPROVAL: AMENDMENT

PRINCIPAL INVESTIGATOR: Anita M. Hubley	DEPARTMENT: Educational & Counselling Psychology, and Special Education	UBC-PHC REB NUMBER: H09-00304

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
Vancouver Coastal Health (VCHRI/VCHA)	Vancouver General Hospital
Vancouver Coastal Health (VCHRI/VCHA)	Lions Gate Hospital
Vancouver Coastal Health (VCHRI/VCHA)	North Shore Community (including Horseshoe Bay and up to Lions Bay)
Providence Health Care	Mount Saint Joseph Hospital
Providence Health Care	St. Paul's Hospital
Vancouver Coastal Health (VCHRI/VCHA)	Vancouver Community

Other locations where the research will be conducted:
Subject's home University of British Columbia Private family physician's office of Dr. Ken Chan, located at 480-2184 W Broadway, Vancouver

CO-INVESTIGATOR(S):
Sherrie L. Myers
Deviani Maher
Martha L. Donnelly

SPONSORING AGENCIES:
British Columbia Network for Aging Research - Michael Smith Foundation for Health Research (MSFHR)
UBC Faculty of Education

PROJECT TITLE:
Criterion, Convergent, and Discriminant Validation of the Hubley Depression Scale for Older Adults (HDS-OA)

REMINDER: The current UBC-PHC REB approval for this study expires: April 13, 2010

AMENDMENT(S):	AMENDMENT APPROVAL DATE: June 22, 2009	
Document Name	Version	Date
Application: The above-stated locations for depressed subject recruitment have been added to the study.		

CERTIFICATION:
1. The membership of the UBC-PHC REB complies with the membership requirements for research ethics boards defined in Part C Division 5 of the Food and Drug Regulations of Canada.
2. The UBC-PHC REB carries out its functions in a manner fully consistent with Good Clinical Practices.
3. The UBC-PHC REB has reviewed and approved the research project named on this Certificate of Approval including any associated consent form and taken the action noted above. This research project is to be conducted by the principal investigator named above at the specified research site(s). This review of the UBC-PHC REB have been documented in writing.

Page 1 of 2

The amendment(s) for the above-named project has been reviewed by the **UBC-PHC Research Ethics Board Chair or Associate Chair** , as presented in the documentation and the accompanying documentation was found to be acceptable on ethical grounds for research involving human subjects.

Approval of the UBC-PHC Research Ethics Board or Associate Chair, verified by the signature of one of the following:

| Dr. Kuo-Hsing Kuo, | Dr. J. Kernahan, | Dr. I. Fedoroff, |
| Chair | Associate Chair | Associate Chair |

Providence Health Care
Institutional Certificate of Final Approval

Principal Investigator:	Department:	Reference Number:
Anita H. Hubley	Educational & Counselling Psychology, and Special Education	PHC H09-00304

Sponsoring Agencies:

British Columbia Network for Aging Research – Michael Smith Foundation for Health Research (MSFHR)

UBC Faculty of Education

Project Title:

Criterion, Convergent, and Discriminant Validation of the Hubley Depression Scale for Older Adults (HDS-OA)

Date Ethical Approval:

April 13, 2009

The UBC-PHC Research Ethics Board granted ethical approval for the above-referenced research project on the date stated above. I am now pleased to inform you that all necessary hospital department/facilities approvals (except SPH Health Records) and institutional agreements/contracts are now in place and that you have permission to begin your research. *

Dr. Yvonne Lefebvre
Vice President Research and Academic Affairs, Providence Health Care
President, Providence Health Care Research Institute

Date: May 13, 2009

* SPH Health Records requires a copy of this certificate prior to granting approval.

St. Paul's Hospital
Holy Family Hospital
Mount St. Joseph's Hospital
St. Vincent's Hospital-Brock Fahrni Pavilion
St. Vincent's Hospital-Langara
Youville Residence

Robert McMaster, D. Phil
Interim
Vice President, Research, Vancouver Coastal Health
Executive Director, Vancouver Coastal Health Research Institute

Associate Dean Research, Faculty of Medicine
The University of British Columbia

April 14, 2009

Dr. Martha Donnelly
Dept. of Psychiatry - Geriatrics

Vancouver Coastal Health Authority Research Study #V09-0086

<u>FINAL CERTIFICATE OF APPROVAL</u>

TITLE: Criterion, Convergent, and Discriminant Validation of the Hubley Depression Scale for Older Adults (HDS-OA)

SPONSOR: British Columbia Network for Aging Research - Michael Smith Foundation for Health Research (MSFHR);UBC Faculty of Education

This is to inform you that your project has been approved and can start immediately. Approval has been granted until **April 13, 2010** based on the following:

1.   UBC-Providence Research Ethics Board Certificate of Approval #H09-00304 (Anita M. Hubley)

2.   VCHA Clinical Trials Administration Office Approval

Yours truly,

For:
Dr. Robert McMaster
Interim Vice-President Research

A joint venture in research between the Vancouver Coastal Health Authority and The University of British Columbia.
Room 100 – 2647 Willow St. Vancouver, BC V5Z 3P1
Tel: 604-875-5641, Fax: 604-875-5684
www.vchri.ca

Vancouver CoastalHealth Research Institute
Healthier lives through discovery

Robert McMaster, D. Phil
Interim
Vice President, Research, Vancouver Coastal Health
Executive Director, Vancouver Coastal Health Research Institute

Associate Dean Research, Faculty of Medicine
The University of British Columbia

July 24, 2009

Dr. Martha Donnelly
Dept. of Psychiatry - Geriatrics

Vancouver Coastal Health Authority Research Study #V09-0086

FINAL CERTIFICATE OF APPROVAL - AMENDMENT**

TITLE: Criterion, Convergent, and Discriminant Validation of the Hubley Depression Scale for Older Adults (HDS-OA)

SPONSOR: British Columbia Network for Aging Research - Michael Smith Foundation for Health Research (MSFHR);UBC Faculty of Education

This is to inform you that your project has been approved and can start immediately.   Approval has been granted until **April 13, 2010** based on the following:

1.   UBC-Providence Research Ethics Board Certificate of Approval Amendment dated June 22, 2009 #H09-00304 (Anita M. Hubley)

2.   VCHA Clinical Trials Administration Office Approval

**** Coastal HSDA (Lions Gate Hospital, North Shore Community) has been added as a study site.**

Yours truly,

For:
Dr. Robert McMaster
Interim Vice-President Research

A joint venture in research between the Vancouver Coastal Health Authority and The University of British Columbia.
Room 100 – 2647 Willow St. Vancouver, BC V5Z 3P1
Tel: 604-875-5641, Fax: 604-875-5684
www.vchri.ca