Subjective Wellbeing Maintenance: Investigating Depression as Suppressed Homeostatically Protected Mood

by

Renée Bittar

BN (Syd), BApp.Sc (Psych) (Hons) (Deakin)

Submitted in partial fulfilment of the requirements for the award of

Doctor of Psychology (Clinical)

Deakin University

November 2009

DEAKIN UNIVERSITY

CANDIDATE DECLARATION

I certify that the thesis entitled:

Subjective Wellbeing Maintenance: Investigating Depression as Suppressed Homeostatically Protected Mood
Submitted for the degree of: Doctor of Psychology (Clinical)

is the result of my own research, except where otherwise acknowledged, and that this thesis in whole or in part has not been submitted for an award, including a higher degree, to any other university or institution.

Full Name _____
(Please Print)

Signed _____

Date .....

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor, Professor Robert Cummins for his unfailing support, guidance, wisdom, expert editing of countless drafts, and for always making me laugh just when I was at the point of despair. Bob, I am forever in your debt.

To my husband who provided me with love and supported me throughout this entire research project, I thank you.

I would also like to thank Ann-marie James for her assistance in the formatting of this thesis.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

GLOSSARY

α H-B	Alpha level calculated according to the Holm-Bonferroni method of correction
%SM	Percentage of Scale Maximum
APA	American Psychiatric Association
AUC	Area Under the Curve
AUWI	Australian Unity Wellbeing Index
BAI	Beck Anxiety Inventory
BDI	Beck Depression Inventory
BDI-II	Beck Depression Inventory – Second Edition
CIDI	Composite International Diagnostic Interview
ComQoL	Comprehensive Quality of Life Scale
DALYs	Disability Adjusted Life Years
DASS	Depression, Anxiety and Stress Scales
DASS-D	Depression, Anxiety and Stress Scales – Depression Subscale
DASS21	Depression, Anxiety and Stress Scales 21-item version (short form)
DASS21-D	Depression, Anxiety and Stress Scales 21-item version – Depression Subscale
DIS	Diagnostic Interview Schedule
DSM	Diagnostic and Statistical Manual of Mental Disorders
DSM-III	Diagnostic Statistical Manual of Mental Disorders – Third Edition
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition – Text Revision
GBD	Global Burden of Disease
GHQ-12	General Health Questionnaire 12-item version
GLS	Global Life Satisfaction
HADS	Hospital Anxiety and Depression Scale
HADS-D	Hospital Anxiety and Depression Scale – Depression Subscale
HPMood	Homeostatically Protected Mood
ICD	International Classification of Disease
ICD-10	International Classification of Disease – Tenth Edition
IDD	Inventory to Diagnose Depression
HRQoL	Health-related Quality of Life

L	Lonngitudinal Survey
LS	Life Satisfaction
MDD	Major Depressive Disorder
MDT	Multiple Discrepancies Theory
PCA	Principal Components Analysis
PCP	Primary Care Physician
PHQ	Patient Health Questionnaire
PWI-A	Personal Wellbeing Index-Adult Version
QoL	Quality of Life
ROC	Receiver Operating Characteristic
SCC	Slope Comparison Calculator
SCID	Structured Clinical Interview for DSM Disorders
SCID-I	Structure Clinical Interview for DSM Axis-I Disorders
SCID-II	Structured Clinical Interview for Axis-II Disorders
SCL-90	Symptom Checklist-90
SCL-90-R	Symptom Checklist-90-Revised
SPSS	Statistical Package for the Social Sciences
SQoL	Subjective Quality of Life
SWB	Subjective Wellbeing
WHO	World Health Organization
YLDs	Years Lived with Disability
Zobs	Observed Value of Z

EXECUTIVE SUMMARY

Over 40 years have passed since the seminal publications of Campbell, Converse and Rodgers (1976) and Andrews and Withey (1976) first drew systematic attention to the area of Subjective Wellbeing (SWB). Despite the accumulation of a vast body of research in this area in the intervening period, debate among investigators over issues of theoretical structure, composition, measurement and definition continues unabated, impeding our understanding of the conceptual characteristics that underlie the interrogation of many of the more difficult questions.

This thesis investigates one of these areas through an exploration of the relationship between SWB and depression. It investigates homeostasis as the mechanism that might underlie the maintenance of SWB, and tests the proposition made within SWB homeostasis theory that depression may be conceptualised as a loss, or more precisely the suppression, of the normal positive sense of Homeostatically Protected Mood (HPMood) subsequent to the failure of homeostasis (Cummins, 2009, in press). This theory posits that SWB is actively managed to lay within a narrow, positive, idiosyncratic 'set-point' range by a system of homeostasis (Cummins & Nistico, 2002). Further, it argues that homeostatic management is actually directed at the protection of HPMood, a major component of SWB.

Presently, there are no data that definitively confirm the operation of homeostatic processes, or verify whether depression is the consequence of failure to maintain HPMood. Given that depression is predicted to be responsible for the largest global burden of disease by the year 2020 (World Health Organization, 2005), if depression may be conceptualised as a loss of SWB, then understanding the mechanisms and processes through which SWB may be maintained would appear critical. In these terms, it is proposed that this research is urgently required and that the investigations herein will assist in achieving these aims.

While it is argued in this thesis that Cummins' propositions are meritorious, the validity of these, and therefore the research undertaken herein, relies on the assumption that SWB and depression are valid in terms of theoretical structure, definition and measurement. However, a review of the SWB and depression literature reveals that many of these fundamental conceptual issues remain contentious. To address these, this research investigates SWB and depression from within the same conceptual framework, that is, from a dimensional perspective. A dimensional understanding of affective experience is one in which mental disorders, such as depression, are seen to exist on a continuum with normality (Eyesenck, Wakefield & Friedman, 1983). Here, Major Depression represents the end-point along a continuum of depressive symptomatology, as opposed to a structurally distinct entity (Ruscio & Ruscio, 2000) This conceptual frame is consistent with the proposition that depression represents the loss of normal positive HPMood, and with the affective framework

within SWB homeostasis theory. Further, it is argued that the use of self-report scales as the reference criteria for depression is more consistent with a dimensional perspective than structured interviews, since these were developed in relation to a categorical taxonomy.

Investigations in this thesis comprise two linked studies. The first study tests a number of the theoretical predictions and diagnostic approximations made within SWB homeostasis theory regarding the relationship between SWB and depression. This has the purpose of testing homeostasis as the proposed mechanism of SWB control. In combination, the results provide general support for the operation of homeostatic processes, and are generally consistent with the notion that depression represents the loss of the normal positive sense of HPMood. Specifically, results confirm that SWB is remarkably stable in the Australian general population. Further, results confirm that SWB demonstrates an inverse curvilinear relationship with depression as predicted by homeostasis theory.

These findings gave rise to a number of additional hypotheses regarding the relationship between SWB and depression. Namely, if depression represents the loss of HPMood, and the underlying structure of depression is dimensional, as found previously (Beach & Amir, 2003), then SWB as a measure of affective normality should evidence utility as an indicator of depression. Testing this hypothesis provides different evidence for Cummins' (2009, in press) theoretical proposition and has important definitional implications. This investigation is undertaken in Study 2.

Study 2 investigates the comparative validity of three widely used self-report depression scales and three measures of SWB as putative, novel measures of depression. These analyses undoubtedly produce the most interesting findings of this thesis. These are that some SWB measures perform as well as some of the most widely used depression scales in the detection of mild depression. In some instances, SWB measures actually outperform depression scales to screen for depression.

Overall, the findings from Study 1 and Study 2 provide support for homeostasis as the proposed mechanism of SWB control, and strong support for the notion that depression represents the loss of the normal positive sense of HPMood subsequent to the failure of homeostasis. Importantly, this research contributes to the growing body of empirical findings that support a dimensional view of affective experience, and in so doing, highlights the need for researchers to consider the congruence between the conceptual frame of their research question and the constructs under examination in any comparative study. Finally, in the context of escalating rates of depression globally, the finding that SWB measures may have greater utility in screening for depression in general population samples is important and requires future research attention.

CHAPTER 1: SUBJECTIVE WELLBEING

Introduction

Subjective Wellbeing (SWB), historically conceptualised as happiness, has been a goal of humankind throughout the ages. Its ubiquitous pursuit may be seen to have evolutionary bases where hedonic or pleasure-seeking drives served to promote *genetic fitness,* or the successful proliferation of one's genes (Darwin, 1871). The composition of happiness, and the most expeditious manner in which it may be attained and maintained over time, has been a topic of philosophical debate and investigation since pre-Socratic times (Tatarkiewicz, 1976).

Resonating closely with SWB, a psychological description of happiness offered within Tatarkiewicz's *Theory of Happiness* (1976), emphasises the subjective, emotional component of the construct, revealing it to be a variant of pleasure based on summing and comparing values and achievements in life, and an absence of negative emotions (unpleasantness). However, the factors that yield and sustain positive emotions may vary between individuals and even within individuals as a function of life circumstances. As such, a precise definition of happiness comprising the same suite of positive emotions that applies to all people across cultures remains elusive. Despite these definitional difficulties, happiness is collectively understood and valued, so much so that in American society, the pursuit of happiness is a right accorded all individuals as proclaimed within their Declaration of Independence (Jefferson, 1776).

Today, the pursuit of happiness holds an honoured position in Western societies and it appears that nearly all people believe, or would like to believe, that they can move constantly in an 'upward spiral' toward ever greater personal wellbeing (Lyubormirsky, Sheldon & Schkade, 2005; Sheldon & Houser-Marko, 2001). This pursuit is not simply a pastime of those who live in relative comfort and safety. Indeed, rather than being an indictment of Western self-centredness, empirical evidence has increasingly demonstrated that the human capacity to be satisfied with life is fundamental to positive mental health and adaptation (Davern, 2004; Davern & Cummins, 2005; Jahoda, 1958; Taylor & Brown, 1988),

fostering numerous beneficial outcomes for societies, communities, families and individuals alike (Fredrickson, 2001; Lyubomirsky, King & Diener, 2004).

For example, individuals with high levels of positive wellbeing have been found to be relatively more pro-social, co-operative, altruistic, and 'other-centred' (Kasser & Ryan, 1996). Additionally, as a group they demonstrate greater productivity, possess higher incomes (Staw, Sutton & Pelled, 1995), have higher rates of marriage and lower rates of divorce, have more friends, richer social interactions (Marks & Fleming, 1999; Okun, Stock, Haring & Wittner, 1984) and exhibit more self-control and greater coping abilities (Aspinwall, 1998; Fredrickson & Joiner, 2002; Lyubormirsky et al., 2005). Further, individuals with high levels of positive wellbeing possess a boosted, more robust immune system (Dillon, Minchoff & Baker, 1985) and perhaps have greater longevity (Ostir, Markides, Black & Goodwin, 2000). All of these factors combined may serve to protect an individual against mental and physical ill health. Given the pervasive impact of SWB, it is not surprising that individuals, organisations, societies and policy makers are interested in how SWB may be cultivated, maintained over time, and its loss prevented.

However, despite over 40 years of research, it is widely acknowledged that the field of SWB remains in its formative stages. Debate among investigators over issues of theoretical structure, composition, measurement and definition continues unabated, retarding our understanding of the conceptual characteristics that underpin the interrogation of many of the more complex questions. One such question relates to the mechanisms underlying the maintenance of individual SWB.

The most comprehensive, parsimonious explanation of the mechanisms that might underlie SWB maintenance is offered by the theory and model of *Subjective Wellbeing Homeostasis* (Cummins & Nistico, 2002; Cummins, 2009, in press). This theory posits that SWB is not free to vary as a function of changing external circumstances, but rather is actively managed to lie within a narrow, positive, idiosyncratic 'set-point' range by a system of homeostasis (Cummins & Nistico, 2002). However, as all homeostatic systems have limits, in the event of a prolonged, strong negative challenge, homeostasis eventually fails, and SWB

levels fall markedly. Cummins (2009, in press) proposes that the failure of the homeostatic system, and the subsequent loss of the normal positive sense of SWB, is the essence of depression.

The proposal that people have a homeostatically controlled set-point for their SWB offers an elegant solution to the complex and elusive issue of SWB maintenance. However, there remains no data to definitively confirm the operation of homeostatic processes, or verify whether depression represents the loss of SWB subsequent to the failure of homeostasis.

In the context of the empirical findings described previously that have demonstrated the significant benefits of positive levels of wellbeing for mental health and adaptation, and in times of rapidly escalating global rates of depression (World Health Organization [WHO], 2005), if depression represents the loss of SWB, then understanding the mechanisms and processes through which SWB may be maintained would appear critical. The research herein investigates the relationship between SWB and depression for the purpose of testing homeostasis as the proposed mechanism of SWB control, and investigating the theoretical proposition that depression may be conceptualised as the loss of SWB subsequent to the failure of homeostasis (Cummins, 2009, in press). First, however, the SWB and depression literature salient to these investigations will be reviewed in three chapters.

This first chapter reviews the studies that led to the creation of a world-wide normative range for SWB. The global consistency of these data will be examined as empirical evidence in support of the theory of SWB homeostasis. Evidence for the short- and long-term stability of SWB will then be presented as further evidence for the proposition of SWB homeostasis. Additionally, methodological issues related to both the normative range and stability data will be highlighted. Then, in order to understand the maintenance mechanism tested in this thesis as potentially responsible for SWB management, the theory of SWB homeostasis (Cummins & Nistico, 2002) will be presented and empirical evidence related to the theory will be reviewed.

In the second chapter, depression is reviewed in terms of its definition, classification, epidemiology, economic costs and measurement. Conceptual issues related to these areas are highlighted and an argument developed for a new conceptualisation of depression in line with a dimensional understanding of affective experience and the theory of SWB homeostasis (Cummins, 2009, in press; Cummins et al., 2009, in press). Additionally, as the interpretability of results from any investigation depend on the validity of instruments used, the three self-report depression instruments used in investigations in this thesis are presented and reviewed. These are the Beck Depression Inventory-Second Edition (BDI-II) (Beck, Steer & Brown, 1996), Depression, Anxiety and Stress Scales (DASS) (Lovibond & Lovibond, 1995), and the Hospital Anxiety and Depression Scale (HADS) (Zigmond & Snaith, 1983).

Finally, the third chapter examines the theoretical propositions and empirical findings related to the relationship between SWB and depression. This review will now commence with the clarification of terminology and measurement issues in the field of SWB and the provision of definitions of basic constructs.

Issues of Terminology and Measurement

SWB research remains relatively nascent and largely atheoretical, with few notable exceptions. Although a taxonomy based on majority opinion is beginning to form, universal agreement on issues of definition, measurement and conceptual framework is yet to be attained. For example, the terms ‘happiness’, ‘quality of life’ and ‘wellbeing’ are still used so inconsistently and at times interchangeably within the one study, that one is forced to search for the author’s interpretation of these terms in order to understand precisely what is being measured. This lack of definitional consistency has affected the operationalisation of constructs and spawned in excess of 600 instruments to measure the SWB of people in a variety of population groupings (Cummins, 2005a). To avoid confusion, only nomenclature and measurement procedures that are judged to be based on current majority opinion within the field (Cummins, 2003, 2005b) will be presented.

The Composition of Subjective Wellbeing

SWB may be defined as a normal state of positive wellbeing and refers to an individual's subjective experience of life as a global entity (Cummins, 2005a). This experience comprises unique cognitive evaluations and affective reactions to personal circumstances (Campbell, Converse & Rogers, 1976) and is influenced by a variety of factors still debated among investigators. These factors include personality, values, expectations, goals, previous experiences (Cummins et al., 2002; Diener, Suh, Lucas & Smith, 1999; Michalos, 1985) and most recently, Core Affect (Davern, Cummins & Stokes, 2007), now termed Homeostatically Protected Mood (HPMood) (Cummins, 2009, in press). HPMood is theoretically described as a biologically determined positive mood that is hard wired for each individual, provides the activation energy for behaviour across the lifespan, and is normally experienced as the combination of contentment, happiness and arousal (Cummins, 2009, in press).

SWB researchers argue that the objective indicators of life circumstances alone do not define *quality of life* (QoL) (Cummins, 1995, 2000a). Individuals react differently to the same circumstances and evaluate conditions based on a unique frame of reference (Diener, Suh, Lucas & Smith, 1999). As such, the term QoL encompasses both objective (e.g., income and health) and subjective components with the terms *subjective quality of life* (SQoL) and *perceived wellbeing*, referring only to the latter.

Most contemporary theorists regard the measurement of SWB to comprise both affect and cognition. The amalgamation of these two processes reflects a combined sensibility that is believed to be best captured through questions of *satisfaction* (Campbell et al., 1976). For example *global life satisfaction* (GLS), representing the most general, abstract, overall rating of *life satisfaction* (LS), is most commonly measured through the classic question 'How satisfied are you with your life as a whole?'. This measure has proved reliable (Larsen, Diener & Emmons, 1985), exhibits moderate stability and appropriate sensitivity to changing life circumstances (Eid & Diener, 2004) and has been found to be particularly consistent in Western countries (Cummins, 1995, 1998, 2003). As such, it has been embraced as an industry standard for examining GLS or global

SWB. It is also widely recognised that GLS may be deconstructed into a number of specific life domains, representing generic life areas that may be both objectively and subjectively described (Cummins, 2005a). This deconstruction is useful to elicit information about the components of life that may contribute positively or negatively to SWB and in aggregate, *domain satisfaction* yields an approximation of GLS (Campbell et al., 1976). While the precise character of these domains is still debated, general consensus has been reached regarding some of the central domains, namely, those involving relationships, wealth and health (Cummins, 1996, 2005a; Diener, Suh, Lucas & Smith, 1999; Headey & Wearing, 1992). As SWB is not a unitary construct, research frequently focuses on either the affective or the cognitive component of SWB.

The affective component of Subjective Wellbeing.

Affect comprises people's moods and emotions and represents their on-line evaluations of their life, personal experiences and life events (Diener et al., 1999). The most general positive affective term is *happiness*, which may be measured by the single question 'How happy are you with your life as a whole?' This term may also be measured through more complex scales that have been traditionally based around the dichotomy of *positive and negative affect* (Bradburn, 1969; Watson, Clark & Tellegen, 1988). Affect is now acknowledged in the literature to comprise two dimensions. One is emotional valence (pleasure or displeasure) and the other is activation (high or low) (Larsen & Diener, 1987; Russell & Carroll, 1999). These two dimensions are depicted as orthogonal axes whose distal poles define a circle in a two-dimensional space (Russell, 1980). This dimensional organisation of affective states forms the basis of the now widely accepted Circumplex Model of Affect (Russell, 1980; Russell & Carroll, 1999). There is no standard scale to measure this conceptualisation of affect at present, although it has been most recently explored by asking participants to 'indicate how each of the following describes your feelings when you think about your life in general' in relation to a list of key affective terms selected to be proportionally representative of terms within the circumplex (Davern, Cummins & Stokes, 2007).

The cognitive component of Subjective Wellbeing.

SWB is based on comparisons of current circumstances with self-imposed standards (Diener, Emmons, Larsen & Griffin, 1985). These standards include other people, past circumstances, ideal circumstances, needs, goals and aspirations (Diener et al., 1999). The idea that satisfaction is based on the discrepancy between current conditions and these multiple standards is best described by Michalos' Multiple Discrepancies Theory (MDT) (1985). Perceptions of current conditions, desired circumstances and self-standards are believed to reflect individual differences, and consequently can vary greatly (Michalos, 1985). As such, discrepancies would appear to be a salient reflection of unique cognitive evaluations of life circumstances. For example, where the comparison standard is judged to be higher than present circumstances, the discrepancy results in decreased satisfaction, whereas if the comparison standard is judged lower than current personal circumstances, this differential results in increased satisfaction.

Researchers have attempted to extract the pure cognitive component of SWB, through the construction of various scales that have employed determinedly cognitive tasks with mixed success (Cantril, 1965). In light of recent studies that demonstrate that SWB is predominantly an affective construct best described in terms of HPMood that perfuses all cognitions (Davern et al., 2007), the creation of a purely cognitive scale may be unachievable.

Measurement Scale Issues

The number of scales purporting to measure SWB or QoL in a variety of contexts and populations is staggering. Many of these, such as Health-Related Quality of Life (HRQoL) scales are designed to measure the SWB of people who are variously compromised (e.g., those living with HIV) and frequently combine objective (e.g., symptoms of disease states) and subjective measures (e.g., patient perceptions of health) to yield a scale whose interpretation is extremely limited. Scales such as these operate on a deficit approach to QoL, such that a high score indicates a relative lack of disease, rather than a high QoL or high SWB. Clearly, these scales bear scant resemblance to SWB as described above. Consequently, it has been proposed that the defining feature of all true SWB scales is the capacity

to yield a score that can be compared to the normative distribution of SWB within the general population (Cummins, 2003). This distribution reflects an empirical frame of reference and provides a normative standard for SWB world-wide. The extensive studies that afforded its derivation and the implications that the establishment of such an empirical standard have for the field of SWB research will now be reviewed and discussed in detail.

Normative Subjective Wellbeing

The overwhelming majority of studies in the field of SWB research have investigated SWB or LS cross-sectionally. Consequently, the proposed normative range of SWB has been based on cross-sectional population sample means. A series of studies have demonstrated that for Western populations, sample means lie within the narrow range of 70-80% of the measurement scale maximum (%SM) (Cummins, 1995, 1998, 2002). In other words, on average, people in Western nations were found to report that they were 70-80% satisfied with their lives as a whole. The statistic %SM was devised by Cummins to convert Likert scale data to a range from zero to100 as a means through which data from a variety of international investigations could be compared.

Normative Range of Western and Non-western Populations

The first of these studies conducted by Cummins (1995) compared seventeen data sets from sixteen unrelated studies employing fourteen variants of global measures of LS. These data sets were selected on stringent criteria and drawn from six Western countries that were considered to possess comparable culture and socio-economic status: Australia, Canada, England, Norway, Sweden and the United States (US). All general population mean scores were recoded, converted to %SM and combined as data to yield a mean and standard deviation of 75 ± 2.5%SM (Cummins, 1995). Thus, two standard deviations around this mean described an approximation of the normative range of 70-80%SM for LS in Western population samples. In a second study, population sample means from 45 countries, representing all geographical regions, were analysed using similar methodology. These data confirmed the population standard of 75 ± 2.5%SM for Western nations. Non-Western nations were found to have a lower margin that

extended below that of Western nations, yielding a World LS mean of 70 ± 5 %SM thereby approximating an international normative range for LS of 60-80%SM (Cummins, 1998). As wealth and individualism, whilst not guaranteeing high population wellbeing, are strong predictors of LS (Cummins, 1998; Diener & Biswas-Diener 2002; Diener, Diener & Diener, 1995; Diener, Sandvik, Seidlitz & Diener, 1993), the inclusion of third-world countries in this study, such as Bangladesh and Kenya, may have contributed to the extension of the normal range lower boundary below that of Western nations. However, there may be other plausible explanations for depressed population LS in collectivist cultures. It is possible that these lower levels of LS in part result from employing LS measurement scales that emphasise the 'I', 'me' and 'mine' form of identity and world view, rather than the 'we', 'us' and 'ours' of a collectivist mentality (Hofstede, 1980), and therefore measure the construct in a non-comparable cultural context. Alternatively, or perhaps additionally, as found by Schimmack and colleagues, the lower rates of LS in Asian populations may result from the propensity of Asian cultures' to view emotions dialectically. This is where emotions of opposite valence are viewed as compatible. As, LS increases with the frequency of pleasant emotions and decreases with the frequency of unpleasant emotions (e.g. Diener, 1984), this way of thinking then influences judgements of LS (Schimmack, Oishi & Diener, 2002).

The coincident upper margin of 80%SM, found in the normal range of LS in both Western and non-Western nations, points to the possibility of ceiling effects. Examination of the pattern of variance within these data confirmed a ceiling effect evidenced by a reduction in variance as sample means increased (Cummins, 1995, 1998). Additionally, floor effects were apparent with even the lowest scoring nations recording a positive mean level of LS (i.e. > 50%SM) (Cummins, 1998, 2003). It has been proposed that the combination of both ceiling and floor effects operate to ensure population wellbeing remains within a positive range (Cummins, 1995, 1998; Headey & Wearing, 1988). The proposed mechanism underlying this pattern of variance and extraordinary consistency of these results is a highly adaptive process of psychological, homeostatic control that functions to maintain LS equilibrium under stable conditions (Cummins,

1995, 1998, 2003; Headey & Wearing, 1988, 1989). This proposition will be examined in detail in Chapter 2.

Within-samples Normative Range

Examination of the distribution of LS within-populations reveals even greater consistency than that displayed by between-population mean LS scores. Data calculated from 62 studies revealed that, in terms of range, the variation in population mean scores was more than four times greater (14.97%SM) than the variation in their corresponding standard deviations (3.49%SM) (Cummins, 2003). The average standard deviation for English-speaking Western nations was 18 ± 1%SM and for all countries was 19 ± 2.5%SM, approximating an intra-population normative range of 40-100%SM and 30-100%SM respectively (Cummins, 2003). Thus, LS was found to be held within a very predictable range within populations, providing further support for the idea that LS might indeed be under homeostatic control.

SWB Values Derived from Three Well-known Indices

Three indices of perceived wellbeing have been systematically investigated and lend varying support for the proposed normative range of 70-80%SM for Western population mean scores. Two of these, the Comprehensive Quality of Life Scale (ComQoL) (Cummins, 1997) and the Personal Wellbeing Index (PWI-A) (Cummins, Eckersley, Pallant, Van Vugt & Misajon, 2003), demonstrate LS distributions in line with the hypothesised normative range (Cummins, 1995). However, the third, the Satisfaction with Life Scale (SLS) (Diener, Emmons, Larson & Griffin, 1985), was deliberately omitted from Cummins' earlier analyses as it was found to produce values of at least 10%SM below other comparable scales (Cummins, 1995). An explanation for these lower values may relate to specific items constituting the scale. Comprising five items on a seven-point Likert-type scale, the inclusion of strong positive statements of LS (e.g., 'the conditions of my life are excellent') may cause people to react strongly to an imposed standard (i.e. 'excellent'). Whereas, scales comprising neutral inquiries (e.g., 'how satisfied are you with your life as a whole?'), encourage participants to use idiosyncratic frames of reference to measure their

satisfaction, possibly preventing strong reactions that may unduly influence responses.

Implications of Empirical Standard for SWB Research

The outstanding consistency of these data has enabled researchers to describe population LS and SWB in terms of an empirical normative standard. The effects of this have been three-fold. Firstly, the interpretability of data from empirical studies on SWB has been enhanced. Prior to the formulation of an empirical frame of reference for SWB, studies were limited by the need to rely on comparisons with other investigations employing identical methodologies. As a result of the extensive lack of agreement on issues of definition and measurement that have plagued the field since its conception, such comparisons were almost impossible. Secondly, it provides a preliminary and integral step towards establishing the adaptive parameters of individual SWB. Thirdly, it has lent support to the idea that SWB possesses a set-point and is held under homeostatic control. This latter proposition is strengthened not only by the predictably narrow normative range of population sample means, but also by the pattern of variability observed within samples. Nevertheless, the normative range for population LS, must be viewed as an estimate only. The method of combining and averaging mean values from studies of widely varying sample sizes must be considered rudimentary. Additionally, despite rigorous study selection criteria, the breadth of methodologies, (e.g., varying instructions to participants, multiple measurement instruments and differing sample composition), employed by the studies comprising the analyses remained vast. In light of the present paper, it is also important to note that consistency in population sample means derived from cross-sectional surveys do not confer stability in individual levels of SWB, and as such, it cannot determine the normative range of individual SWB. The evidence relating to this concept of temporal stability will now be reviewed in detail.

Subjective Wellbeing Stability

Experimental studies have revealed that SWB judgements may be vulnerable to occasion-specific influences. In a series of experiments, Schwarz and Strack (1991, 1999) demonstrated that self-reports of LS can be strongly

influenced by situational conditions, such as state mood, eliciting only small test-retest correlations for LS even despite short test-retest intervals. These researchers hypothesised that this effect should be stronger for reports of global SWB (i.e. satisfaction with 'life as a whole') than for LS measures with well circumscribed evaluative criteria (i.e. satisfaction with specific life domains), possibly due to the higher level of abstraction inherent in global SWB measures. Further, Schwarz and Strack argued that unless the informational value of an individual's affective state was questioned, momentary mood was used as an indicator of wellbeing. As such, the validity of conceptualising SWB as exhibiting relatively stable, trait-like properties, even over the short-term, has been questioned.

One method currently employed by researchers to overcome this problem has been the investigation of SWB over time using non-experimental, longitudinal research designs. However, to date there has been a comparative paucity of research investigating SWB in this manner. There may be several reasons for this. Longitudinal, relative to cross-sectional research, is more time consuming and more difficult to administer. The data collected are more difficult to maintain, and cumulative attrition rates in general population samples are high. Additionally, the field is in its infancy, with its development stifled by a lack of agreement between investigators regarding such fundamental issues as the definition and composition of the central construct. Consequently, examination of complex concepts requiring longitudinal investigation has not been as highly prioritised as cross-sectional research in the past.

Long-term Stability

The majority of extant non-experimental, longitudinal evidence demonstrates moderate to substantial support for the stability of SWB over time. For example, Costa and McCrae (1989) reported correlations of 0.47 to 0.63 using multiple measures of SWB over a two-year period. Similarly, using their Life Satisfaction Index in four waves of a general population panel study, Headey and Wearing (1989) reported test-retest correlation coefficients for LS of 0.64, 0.51 and 0.52 at two, four and six-year intervals respectively. Two studies of elderly individuals have reported correlations in the range of 0.46 to 0.66 for LS over a three-year period (Baur & Oken, 1983; Bowling, Farquar & Grundy, 1996).

Further, the affective component of SWB, frequently measured through reported levels of positive and negative affect and through measures of emotional intensity, has also been found to be moderately stable. In student samples, Watson and Walker (1996) reported six to seven-year stability coefficients for positive and negative affect in the range of 0.36 to 0.46, and Costa and McCrae (1988) found six-year stability coefficients in the 0.50 range for both positive and negative affect in an adult sample.

Short-term Stability

Relative to long-term stability coefficients, test-retest correlations for short-term SWB stability are larger, whereas state mood demonstrates much greater variability over the short-term. Over time intervals of three and twelve weeks, Steyer, Ferring and Schmitt (1992) found relatively small test-retest correlations for state mood, of between 0.22 and 0.52. In comparison, test-retest correlations for global judgements of SWB were much larger (r = 0.82, Pavot & Diener, 1993). For emotional intensity across time intervals of one to three months, test-retest correlations were found to be between 0.81 and 0.83 (Bacharowski & Braaten, 1994; Larsen & Diener, 1987).

Using an alternative approach, Diener and Larsen (1984) demonstrated a high level of support for the stability of global and affective SWB between multiple daily measures of LS and positive and negative affect. They proposed that if SWB reflected an individual's evaluations and affective reactions to 'life as a whole', then contrary to Schwarz and Strack's (1991, 1999) hypothesis, self-reported SWB should not be determined by situational factors, such as state mood, evidenced by short-term stability across varied situations. In order to test this proposition, Diener and Larsen (1984) measured LS, positive affect and negative affect by asking participants to complete self-report inventories multiple times a day for multiple days across dissimilar situations. These situations included work *versus* recreation, novel *versus* typical, and alone *versus* social situations. They found that in diverse situations LS in work situations correlated 0.95 with LS in recreation situations and similar correlation coefficients were reported across all other types of situations.

Methodological Issues and Subjective Wellbeing Stability

These results, supportive of SWB stability, cannot not be seen to reflect the true stability of SWB as measurement error was not considered. Short-term stability was estimated by correlations of observed variables and, therefore, measurement error and instability were confounded. Consequently, if measurement error is not accounted for in longitudinal studies, true stability will be under-estimated (Eid & Diener, 2004). Analyses that are more sophisticated have since addressed this issue and confirmed that the true stability of global SWB is very large (Deinzer, Steyer & Eid et al., 1995; Eid & Diener, 2004; Ferring, Filipp & Schmidt, 1996). Taking measurement error into account, two studies have demonstrated that for LS, only 7-18% of the variance on an occasion of measurement was due to occasion-specific variability (Deinzer et al., 1995; Ferring et al., 1996). However, in these studies, the occasion-specific fluctuations had not been related to state mood.

Taking random error, occasion-specific deviations and stable inter-individual differences into account in a multistate-multitrait-multiconstruct-model, Eid and Diener's (2004) most recent analyses confirmed that self-reported LS exhibited trait-like properties. In this study, LS, frequency and intensity of emotions, and personality ratings (self-esteem, optimism, neuroticism and extraversion) were assessed in a college student sample on three occasions with intervals of four weeks between measurements. This time interval was chosen to ensure assessments reflected the short-term variability of LS, rather than important changes in life circumstances, while remaining sufficiently long to prevent any carry-over measurement effects (Eid & Diener, 2004). On the three occasions of measurement, substantial stability was found in LS with 74-83% of the variance explained by stable inter-individual differences, 12-16% of the variance due to occasion-specific influences (occasion-specificity) and 4-10% due to random error. Importantly, on two occasions of measurement, the occasion-specificity correlations between state mood and LS were not significantly different from zero (r = 0.13, r = 0.23), indicating a very weak association between state mood and LS. However, for the state mood scales the occasion-specificity coefficients were comparatively large, being greater than the

consistency coefficients on two occasions of measurement. Combined, these results indicated that variability in SWB was very small, that occasion-specific influences in self-reported LS were also small, and that the association between state mood and SWB was relatively weak.

In summary, these studies indicate that SWB is relatively stable over short and long periods and in diverse contexts. SWB is not significantly influenced by situational factors, such as momentary mood, under non-experimental conditions. However, methodological limitations must be acknowledged, in particular that stability correlations reviewed in this section are based on sample means and as such, cannot infer temporal SWB stability at the level of the individual. Nevertheless, in conjunction with the consistent and predictable normative ranges of SWB, these data constitute accumulating evidence to support the proposition that SWB is not free to vary as a function of personal circumstances, but rather, is actively controlled and maintained over time through homeostatic processes.

The Maintenance of Subjective Wellbeing

Consistent with the idea of homeostasis, the earliest data on positive mental states revealed that individuals' reported levels of happiness lie naturally within a positive range (Goldings, 1954). In 1976, two seminal publications provided compelling evidence that global SWB, and its related constructs, exhibit trait-like properties that could be measured predictably and reliably over time (Andrews & Withey, 1976; Campbell et al., 1976). Within these publications, two of the earliest models proposed responsible for SWB were described. These were the 'Lewininan Lifespace Model' (Campbell et al., 1976) and the 'Two-Dimensional Conceptual Model' (Andrews and Withey, 1976) and were primarily concerned with the separation of QoL into its objective and subjective components. However, more than a decade passed before the researchers incorporated the properties of stability and determined positivity into their SWB models.

The first of these more recent models was Heady and Wearing's (1986, 1987, 1989) 'Dynamic Equilibrium Model'. Using data from a panel study, these researchers observed that people appeared to possess 'normal' equilibrium levels

of SWB, or a ‘set-point,’ for their SWB. Specifically, Heady and Wearing (1992) observed that a positive or negative life event may enhance or depress SWB respectively, yet over time, people had a propensity to return to their pre-event levels of SWB.

More recently, Headey and Wearing (1989, 1992) proposed a Dynamic Equilibrium Model in which each individual was regarded as possessing ‘normal’ equilibrium levels of SWB, or a set-point for LS. These normal levels were believed stable unless disrupted by a positive or negative life event that may enhance or depress SWB respectively. However, these researchers observed that over time people had a propensity to return to their pre-event levels of SWB (Headey & Wearing, 1992).

The proposed regularity of normal SWB was corroborated and expanded by a series of investigations conducted by Cummins (1995, 1998, 2003). These investigations resulted in an empirical frame of reference for normative SWB. As previously discussed, these studies demonstrated that Western, non-Western and world population mean levels of SWB predictably lie within consistent, narrow ranges that are normally positive (Cummins, 1995, 1998, 2003). That is, on a scale of zero to 100, where zero represents complete dissatisfaction with life and 100 represents complete satisfaction, in response to some variant of the classic question ‘how satisfied are you with your life as a whole?’, individuals in Western nations report SWB levels that lie within the positive (50-100) or *satisfied* sector of the *dissatisfied-satisfied* continuum (Cummins, 2003). In order to explain this narrow, positive range of values, Cummins hypothesised that SWB is regulated actively and internally. This proposition forms the basis of the most comprehensive theoretical and explanatory model of SWB control to date; the theory and model of *Subjective Wellbeing Homeostasis* (Cummins & Nistico, 2002).

An overview of the theory of SWB homeostasis will precede a review of the evidence relevant to the testing of homeostasis as the proposed mechanism of SWB control. The general predictions regarding the behaviour of SWB to meet the standard performance requirements of homeostatic systems will be summarised. Following this, HPMood (Cummins, 2009, in press) as the major

component of SWB will be defined and explained and the breakthrough studies that have demonstrated that SWB is predominantly an affective construct best described in terms of HPMood are reviewed. However, the precise predictions regarding the changing levels of SWB as homeostasis is challenged that are relevant to the relationship between SWB and depression are not presented here. These will be presented in Chapter 3.

Subjective Wellbeing Homeostasis Theory

The term homeostasis was coined from two Greek words *homoios* meaning *same* and *stasis* meaning *steady* (Cannon, 1932) and may be defined as 'the maintenance of equilibrium in any physiological, psychological, or social process by an automatic feedback mechanism compensating for disrupting changes' (Colman, 2003, p. 335). The theory of SWB homeostasis has incorporated these basic principles as a means through which the mechanisms of SWB maintenance may be explained.

Homeostasis theory posits that, analogous to the homeostatic regulation of body temperature and electrolyte balance, SWB is actively managed and maintained within an idiosyncratic set-point range by a suite of psychological devices that have evolved for this purpose (Cummins & Nistico, 2002; Cummins, Gullone & Lau, 2002). Further, it is proposed that this management is actually directed at the protection of HPMood, which we normally experience as a combination of contentment, happiness, and arousal (Davern, 2004).

Under normal operating conditions, HPMood defines the set-point at which all other devices implicated in the homeostatic maintenance of SWB are held (Davern & Cummins, 2005). HPMood as the proposed set-point, and major constituent of SWB, will be explained in detail in this chapter. In the context of Cummin's (2000a, 2003, 2005a) homeostatic model, these psychological devices implicated in the homeostatic process and intrinsically linked to levels of SWB (De Neve, 1999), comprise perceived control, self-esteem and optimism. As such, HPMood functions as a parsimonious indictor of general, personal wellbeing at these times.

Also incorporated in this theory is the concept of a critical 'threshold', believed to exist at the boundaries of an individual's set-point range. In line with the principles of homeostatic systems, Cummins and colleagues propose that in response to a challenging agent (e.g., negative life event), the homeostatic system resists change in an attempt to maintain SWB within the parameters of the set-point range. As SWB approaches the threshold for homeostasis, the constellation of psychological devices operates as a cognitive buffering system to oppose further change.

However, all homeostatic systems have their limits. According to theory, a sufficiently adverse environment or event can defeat the homeostatic system and when this occurs, the level of SWB falls below the threshold of the set-point range. Cummins (2009, in press) proposes that the loss of the normal positive sense of HPMood subsequent to the failure of homeostasis is the essence of depression. This proposition will be explored in Chapter 3. Nevertheless, it is important to note that several studies have found general support for this proposition. For example, families who experience the chronic stress of caring for young people with intellectual disabilities at home had significantly lower SQoL, particularly in the domains of health and productivity, than those without offspring with an intellectual disability (Browne & Bramston, 1996). These resilts were corroborated by Cummins (2001), who found that people who care for a severely disabled family member at home have low levels of SWB and most importantly, these people as a group (carers) have been found to also have higher than normal levels of depression (Cummins, 2001; Spangenberg & Theron, 1999; Stuckey, Neundorfer & Smyth, 1996; Tymchuk, 1994; Whitlatch, Feinberg & Sebasta, 1997).

However, over time and in the absence of an overwhelming negative force, the homeostatic system functions to regain control and restore levels of SWB to its set-point (Hanestad & Albreksten, 1992; Headey & Wearing, 1989; Suh, Diener & Fujita, 1996). Hence, homeostasis theory predicts that, under normal operating conditions, levels of SWB will show little relationship to normal variations in an individual's chronic, objective circumstances of living (Cummins, 2000a; Cummins et al., 2002). This is because homeostasis, rather than external

conditions, is controlling levels of SWB. This lack of relationship between SWB levels and external circumstances under maintenance conditions has been confirmed in an empirical review (Cummins, 2000b).

From this description, and according to Cummins (2009, in press), describing SWB maintenance in terms of homeostasis makes a number of very clear predictions regarding the relationship between SWB and other variables. Specifically, Cummins and colleagues (2009, in press) propose that SWB must conform to the standard performance requirements of homeostatic systems and display the following characteristics:

1. SWB must be highly stable
2. There must be a threshold value that is defended by the homeostatic processes. There must also be evidence that as this value is approached the system works harder than normal to retain control. Then, as the threshold value is exceeded, there must be evidence that homeostasis has failed and is no longer controlling the level of SWB.
3. Following homeostatic defeat, over time the system should act to regain control. If this is successful, the level of SWB should return to a stable approximation of its set-point.
4. The aim of homeostasis is to maintain the variable it is managing within a narrow range of values. Thus, SWB must evidence a 'set-point range' that reflects the normal moment-to-moment range in which SWB will be found for each individual. The magnitude of this range may also be an individual difference, with some ranges being more tightly controlled than others are.
5. SWB should respond to variables that either enhance or challenge the operation of the homeostatic system. However, the nature of the relationship with such variables should be consistent with the operation of a homeostatic system.

The implications of these requirements for the investigations of SWB and depression conducted in this thesis are elaborated in Chapter 3.

Evidence for Subjective Wellbeing homeostasis.

As previously described, empirical evidence for SWB homeostasis exists at the level of population sample means. Detailed analyses of within-populations sample variance for GLS in Western nations revealed a systematic increase in variance as sample means fell below the hypothesised normative lower range threshold of 70%SM (Cummins, 2003). This indicates that while SWB is maintained within the homeostatic range of 70-80%SM, it is being held quite stable. However, once the homeostatic system is challenged, homeostasis theory predicts that decreasing levels of SWB are associated with increasing instability as SWB falls under the control of the challenging agent (Cummins et al., 2002). It is proposed that at these times, the relative strength of the relationship between levels of SWB and the aversive agent or event is enhanced. Cummins (2000b) found a generally higher correlation between SWB and objective variables under conditions of increased extrinsic threat specify (e.g., poverty), providing preliminary support for this changing relationship between wellbeing and objective circumstances.

Relative to the increased instability in global LS mean scores below the normative lower threshold of 70%SM, increased stability in mean scores above 70%SM demonstrated a maximum upper threshold of 80%SM. Verification for this upper normative threshold has been confirmed by a number of extensive studies finding no samples significantly exceeding this normative range upper boundary (Cummins, 1998, 2003). These studies have examined mean LS values in sub-groups known to demonstrate some of the highest group levels of perceived wellbeing, namely, the very wealthy (Groot & Vandenbrink, 2000), Nordic populations (Cummins, 1995) and American ‘Back-to-Landers’ (Jacob & Brinkenhoff, 1999). They found that none of these groups possess mean LS values significantly above the upper normative range boundary. Thus, 80%SM would appear to be the maximum LS value that can be supported by a representative general population sample, giving credence to the idea that SWB is not only actively managed within narrow, adaptive parameters, but that there appears to be an average genetic ceiling for individual SWB.

Consistent with the principles of homeostatic systems, homeostasis theory predicts that people who suffer an episode of homeostatic failure, which results in the suppression of normal levels of SWB, should improve or regain their previous levels of SWB over time (Cummins, 1998; Cummins et al., 2002). This has been widely reported, for example, in people with Type 1 diabetes (Hanestad & Albreksten, 1992), and in people who experience SWB loss due to a wide variety of life events of varying intensity, ranging from 'daily hassles', to 'death of a friend'(Headey & Wearing, 1989; Suh, Diener & Fujita,1996). Additionally, a wide variety of studies tracking the SQoL of individuals after having been exposed to events that severely impact their health provide support for the recovery of SWB levels over time. Examples include: people with severe burns (Andreasen & Norris, 1972; Patterson, Everett, Bombardier, Quested, Lee & Marvin, 1993); people who had undergone cardiac surgery (King, 2000; Lindquist et al., 2003); people diagnosed with cancer (Bloom, Fobair, Spiegel, Cox, Varghese & Hoppe, 1991); and people who had spinal cord injuries (Bach & Tilton, 1994). However, methodological issues limit the utility of such data. Specifically, a baseline SWB measurement for each individual against which the extent of loss and recovery of SWB could be quantified is absent. Therefore, in the absence of a known set-point for each individual, it is uncertain whether recovery of SWB is complete or incomplete if an individual improved their SWB over time from 50 to 60%SM. In these terms, an important question yet to be resolved is how SWB improvement may be best measured in individuals with SWB set-points that exist below the normative population range of 70-80 points (Cummins, 1995, 1998).

Hence, while some evidence exists to demonstrate SWB recovery over time, the precise strength of an event required to cause SWB loss, and the variable degree of recovery shown by some individuals exposed to the same event is unknown. Similarly, if the set-point for SWB is individually defined, then homeostatic failure may have little to do with the extent to which an individual deviates from the mean range of a given population. Rather, susceptibility to homeostatic failure may depend on the propensity of an individual to react intensely to conditions of challenge and fluctuate widely outside the parameters of their set-point range. Therefore, the critical threshold at which homeostasis fails

may vary considerably between people, as it is individually defined. In these terms, Cummins' (2003) proposition that normal, idiosyncratic ranges of SWB lie between 50-100%SM appears reasonable as, presumably, only positive (>50%SM) SWB is adaptive.

Nevertheless, the proposition of an absolute threshold for homeostasis appears an empirical question. If the set-point range for SWB is individually defined as Cummins and colleagues suggest, then it may be possible that an individual possessing a normal set-point range of 65-75 points may be depressed at 60 points, since for this person, the critical threshold of their homeostatically controlled set-point has been breached. In this context, if depression represents the loss of the normal positive sense of HPMood, as Cummins proposes (2009, in press), it is possible that depression will present at various levels of positive wellbeing. This proposition will be investigated in Study 1 in this thesis.

Of fundamental importance to any investigation of SWB homeostasis theory, and of particular relevance to the relationship between SWB and depression, is an understanding of the major constituent of and set-point for SWB. This review will now examine the evolution of HPMood as SWB set-point as proposed within homeostasis theory.

Homeostatically Protected Mood

From Personality to Core Affect.

Integral to the theory of SWB homeostasis is the concept of a set-point. That is, each individual is proposed to have an innate set-point for their normal level of SWB, around which their perceived wellbeing is usually held within a narrow range (set-point range). Until most recently, personality was believed responsible for the affective component of SWB and for the determination of the set-point range at which all other systems implicated in SWB maintenance is held (Cummins et al., 2002). This belief was based on a vast body of empirical literature yielding strong converging evidence for personality's involvement in reported levels and stability of SWB (Brief, Butcher, George & Link, 1993; Doyle & Youn, 2000; Kozma, Stone & Stones, 1997). This involvement was most particularly related to the dimensions of extraversion and neuroticism. Whereas

extraversion was found to be strongly associated with positive affect and positively correlated with SWB, neuroticism demonstrated a strong association with negative affect and was inversely related to SWB (Costa & McCrae, 1980; 1984; De Neve & Cooper, 1998, Lucas, Diener, Grob, Suh & Shao, 1998; Vitterso, 2001; Vitterso & Nilsen, 2002).

Recent research (Davern, 2004; Davern et al., 2007) has debunked the notion that SWB is best predicted by personality and contradicted the idea that SWB represents a dominantly cognitive evaluation as proposed by Diener, Napa, Scollon and Lucas (2004). Rather, these recent studies have suggested that the fundamental nature of SWB is best described in terms of Russell's (2003) concept 'Core Affect' (now termed HPMood by Cummins, 2009, in press).

According to Russell (2003), Core Affect is a neurophysiological state underlying the conscious experience that may be conceptualised as a deep form of mood, or trait affect. Russell proposed that Core Affect manifests as a unitary blend of hedonic (pleasure-displeasure) and arousal (activation-deactivation) values that may, at any given moment, be plotted as a single point on the affective circumplex depicted in Figure 1. Further, Russell proposed that Core Affect is subjectively experienced, omnipresent, exists without being attributed to any cause (object-free), and can be consciously accessed when attention is drawn to it. Thus, Core Affect is a biologically influenced mood and refers to how the individual generally experiences himself or herself in a personal but abstract way.

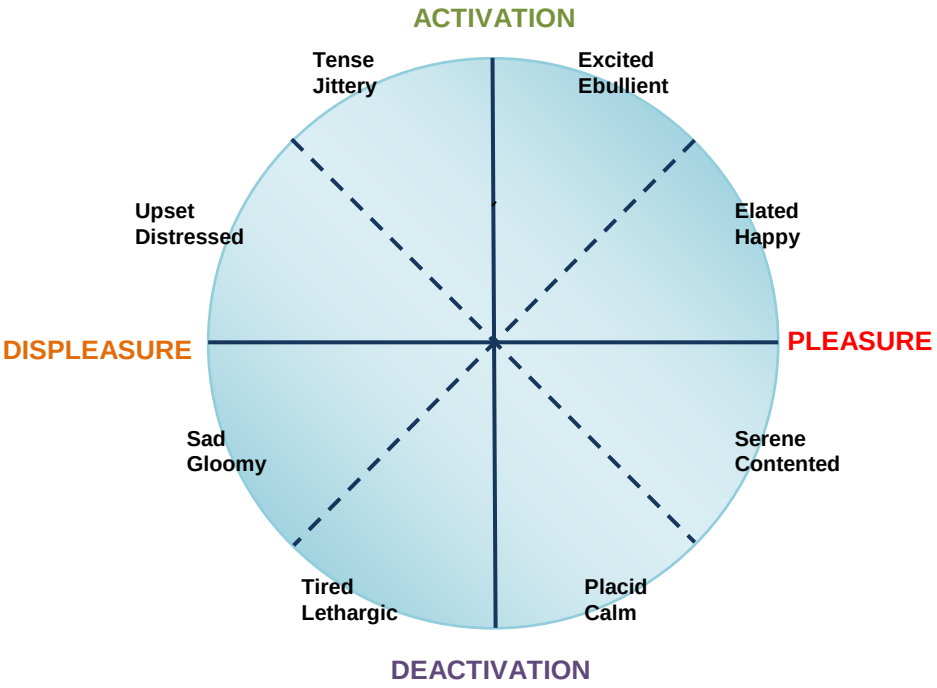


Figure 1. Circumplex Model of Affect (Russell, 1980, 1999): A diagrammatic representation of a spatial model capable of accounting for all affective states.

In the recent studies conducted by Davern et al. (2007), the relative strength of affect, cognition, and all five factors of personality as predictors of SWB was examined in a sample of 854 general population Australian adults. On the basis of earlier research by Davern (2004) the affective component of SWB was operationalised to measure Core Affect using three affect items (happy, content, excited) that represented the activated and deactivated pleasant quadrants of the affective circumplex depicted in Figure 1. The cognitive component of SWB was measured using seven items derived from MDT (Michalos, 1985) described previously. These items addressed the perceived gap between what the respondent currently has and general life aspirations, what age-matched others have, the best one has had in the past, expected to have three years ago and expects to have after five years, deserves and needs.

Davern et al. (2007) found that all three components correlated significantly with SWB and with one another. However, when the variances were controlled by structural equation modelling, it was demonstrated that affect and MDT (Michaelos, 1985) were the dominant components of SWB. Importantly, these researchers found that after accounting for both affect and MDT, personality

accounted for only a very small amount of variance in SWB. A simplified version of their model is reproduced in Figure 2.

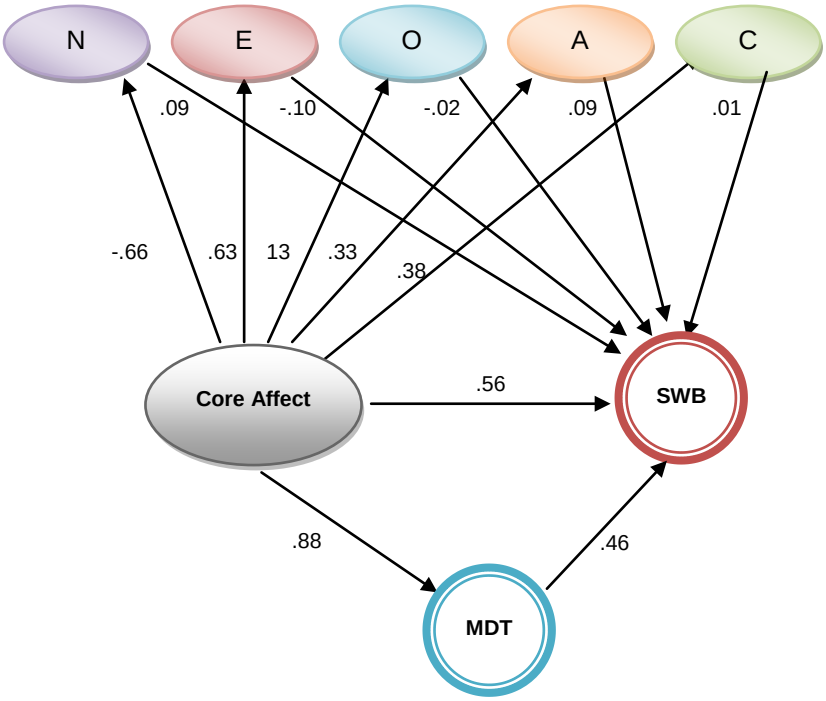


Figure 2. Simplified affective-cognitive model of SWB (Davern, Cummins & Stokes, 2007).

In Figure 2, MDT (Michaelos, 1985) is designated as MDT and the personality factors are designated as: N-Neuroticism, E-Extraversion, O-Openness, A-Agreeableness, C-Conscientiousness. From these results, Davern et al. (2007) concluded that the dominant component of SWB is mood that may be best described in terms of Russell's (2003) concept Core Affect. Importantly, Davern et al.'s (2007) findings have been replicated twice, using independent data (Blore, 2008; Tomyn, 2008). In combination, these results provide strong support for the proposition that SWB is essentially an affective construct.

From Core Affect to Homeostatically Protected Mood.

Since the publication of Davern et al.'s (2007) findings, Russell (2009) has moved away from his description of Core Affect as an object-free biologically influenced mood. Specifically, Russell has revised his definition of Core Affect

and now proposes that Core Affect may be involved in either moods or emotions, that it may even be changed by various other influences, and that it may become directed at or attached to something. In response to this revision, Cummins (2009, in press) has deemed this most recent description of Core Affect incompatible with Russell's (2003) earlier definition of the construct, and contrary to the meaning implied by Davern et al. (2007) when using the term in relation to SWB and homeostasis. Consequently, Cummins proposed a new term, HPMood, to describe the deep trait form of mood associated with SWB and homeostasis theory in keeping with Russell's earlier conceptualisation of Core Affect. Cummins (2009, in press) describes HPMood as a feeling state with the following characteristics:

1. HPMood is a biologically determined positive mood that comprises the most basic experienced feeling. It is hard-wired for each individual, comprising the tonic state of affect that provides the activation energy, or motivation, for behaviour.
2. HPMood is not only the dominant affective constituent of SWB, as determined by Davern et al., but also the basic steady-state, set-point that homeostasis seeks to defend.
3. HPMood perfuses all higher process, including personality (for a review of the neurobiology of personality see Depue & Collins, 1999), memory and momentary experience. It perfuses all cognitive processes to some degree, but most strongly the rather abstract notions of self (e.g., I am a good person). These self-perceptions are held at strength of positivity that approximates the set-point HPMood.

In addition to the above characteristics, Cummins (2009, in press) proposes that a level of HPMood corresponding to 75 points on average, or within the range 70-80 points pleasant or positive comprises the optimum set-point range for SWB, as this corresponds to the most adaptive range for mood. The reason offered by Cummins is that both higher and lower levels of SWB are each associated with cognitive advantages and disadvantages and therefore a range of

70-80 points positive represents a trade-off between the advantages and disadvantages of these higher and lower levels of SWB.

Chapter 1 Conclusions

A number of conclusions can be drawn from the present analysis of the literature as it relates to SWB and homeostasis theory. Based on the data pertaining to the normative distribution of LS and pattern of variance observed within-population samples, the proposition that SWB is not free to vary over its theoretical range is supported. The moderate stability of the construct over time and across diverse situational contexts provides further support for this proposition. However, the proposed normative range for SWB has been based on population sample means (Cummins, 1995, 1998, 2002) and evidence for SWB stability has also been generated from sample mean scores. Consequently, these data do not confer stability at the level of the individual. Notwithstanding this, the remarkable stability of SWB at the population level affords an opportunity to explore the operation of homeostasis processes at a group level. This investigation will be undertaken in Study 1 in this thesis.

The examination of SWB homeostasis theory illuminated the interconnected psychological devices comprising the proposed homeostatic management system for SWB. In addition to the findings regarding SWB stability, empirical studies on the recovery of SWB levels following traumatic events have provided a range of evidence that in combination provide general support for homeostasis as the proposed mechanism of control. The finding that SWB is essentially an affective construct that may be best described in terms of HPMood has been replicated using independent data on two subsequent occasions. Together, these findings provide strong support for the theoretical proposition that SWB management is actually directed at the protection of HPMood and are consistent with the proposition that HPMood defines SWB set-point.

Based on the predictions made within homeostasis theory regarding the required behaviour of variables under homeostatic control, there is an opportunity for this research to test homeostasis more precisely as the proposed mechanism of

SWB control by testing these predictions. Additionally, as these predictions require that SWB be examined in the context of another or other variables, there is an opportunity to simultaneously test homeostasis as the controller of SWB and investigate the proposition that depression represents the loss of normal levels of SWB subsequent to the failure of homeostasis through examination of the relationship between SWB and depression. These investigations will be undertaken in two linked studies in this thesis. This review will now turn to the examination of depression.

CHAPTER 2: DEPRESSION

Clinical depression is one of the most common psychiatric disorders, with its pervasive impact felt from the level of the affected individual to broader society. For the individual, depression is debilitating, results in considerable distress, causes impairments in social and occupational functioning, and represents significant loss of life quality (American Psychiatric Association, 2000). In severe cases, depression may require frequent hospitalisation and can lead to suicide, with depressive disorders found in 30-90% of all completed suicides world-wide (Lönquvist, 2000). Projections from epidemiological data and health economic measures indicate that by 2020, depression will be responsible for the largest global burden of disease (GBD) (Murray & Lopez, 1996; WHO, 2001).

In this chapter, depression is discussed in terms of its definition, classification, epidemiology, economic costs and measurement. Throughout this chapter, conceptual characteristics of the construct are highlighted and an argument is developed for a new conceptualisation of depression that is consistent with a dimensional framework for understanding and measuring affect. Additionally, the three self-report instruments germane to the studies in this thesis are presented and reviewed. These are the BDI-II (Beck, Steer & Brown, 1996), DASS (Lovibond & Lovibond, 1995), and the HADS (Zigmond & Snaith, 1983).

Issues of Definition

A distinction must be made between the meaning attributed to the term ‘depressed’ by lay people, and the meaning of depression in psychology and medicine. For lay people, the term ‘depressed’ is frequently used as a synonym of words such as ‘upset’, ‘sad’ or ‘low’ to describe transient affective states (Faravelli, Ravaldi & Truglia, 2005) . This is similar to the American Psychological Association’s (2007) definition of the term, where depression is defined as ‘dysphoria that can vary in severity from a fluctuation in normal mood to an extreme feeling of sadness, pessimism, and despondency’ (p. 269). In psychology, this meaning may be implied when describing a client’s affect or mood as ‘depressed’ on assessment (Morrison, 1995).

In contrast to the above view of depression as a continuum of depressed mood, the concept of depression as a disorder is primarily medical. Here it is viewed as a syndrome that differs from normal feelings of sadness in the following ways: it is quantitatively and qualitatively more severe; it is not influenced by reality, external situations or events; its duration is longer; and it gives rise to a suite of other symptoms (American Psychiatric Association, 2000; Faravelli et al., 2005). Within this medical framework, the presence of depressed mood or anhedonia, although central to the idea of depression as a disorder, is not sufficient for diagnostic purposes (American Psychiatric Association, 2000). In this medical context, it follows that irrespective of the self-reported or clinically judged severity of depressed mood, no diagnosis of a depression as a disorder would be made based on depressed mood alone. Therefore, in medicine, depression refers to a broader psychobiological construct comprising an assortment of affective, cognitive (e.g., indecisiveness), and somatic symptoms (e.g., appetite disturbance) that may manifest to varying degrees, and includes depressed mood or anhedonia. In combination, these groups of symptoms comprise the criteria used to diagnose depression within the two official taxonomies for common mental disorders. They are the Diagnostic and Statistical Manual of Mental Disorders (DSM), the current version of this manual being the Fourth Edition-Text Revision (DSM-IV-TR; American Psychiatric Association, 2000), and the International Classification of Disease (ICD), with the Tenth Edition (ICD-10; WHO, 1992) being the current version.

Classification of Depression

The DSM-IV-TR (American Psychiatric Association, 2000) and ICD-10 (WHO, 1992) adopt a disease approach to classifying psychological disturbance centring on the notion of syndromes that form exclusive categories within a categorical framework. Here disorders are viewed as discrete entities that occur independently of other discrete disorders, although they can coexist and give rise to comorbidity (Ingram & Siegle, 2002). Additionally, depression is viewed as a mental disorder, or mental illness, rather than as a psychological or behavioural problem. Nevertheless, in the ‘Preface’ of the DSM-IV-TR, it is noted that mental disorders are not necessarily *true* categorical variables. Specifically, the:

> DSM-IV-TR is a categorical classification that divides mental disorders into types based on criteria sets with defining features ... [T]here is no assumption that each category of mental disorder is a completely discrete entity with absolute boundaries dividing it from other mental disorders or from no mental disorder. There is also no assumption that all individuals described as having the same mental disorder are alike in all important ways (p. xxxi).

Despite this qualification, the DSM classification system constitutes a categorical framework for understanding psychological distress and is similar to that used to identify and diagnose physical disorders. Consequently, the utility of a categorical approach to classifying mental disorders that exist on a continuum with normality (Mirowsky, 1994; Ruscio & Ruscio, 2000; Vrendenberg, Flett & Krames, 1993; Widiger, 1997), as opposed to within discrete categories (Goodwin & Guze, 1989; Guze, 1992; Robins & Helzer, 1986), has been debated for almost one hundred years.

Critics of the categorical approach have argued that dimensional systems may better reflect the true nature of many forms of psychopathology, convey more information, and have greater reliability (Eysenck, Wakefield & Friedman, 1983; Flett, Vrendenberg & Krames, 1997; Shankman & Klein, 2002). These researchers argue that diagnoses should take the form of continuous ratings along one or more dimensions. The proponents of a categorical taxonomy dispute this, arguing that mental disorders should be diagnosed as either present or absent. With respect to the classification of depression, the dimensional *versus* categorical debate has recently focussed on whether subclinical/subthreshold forms of depression lie on a continuum with clinical cases (Judd, 1997; Pincus, Davis & McQueen, 1999) or whether these represent different classes (Solomon, Haaga & Arnow, 2001). Consequently, whether subclinical depressive states can be used as indicators of clinical depression risk, or as substitutes to further understanding of the aetiology of clinical depression (Weary, Edwards, Jacobson, 1995), has important implications for the study, assessment and treatment of depression. To date, this debate remains unresolved. Resolution is hindered by methodological issues related to differences in sampling (clinical *versus* student samples) and

assessment techniques (self-report inventories *versus* clinical interviews) used by researchers in the field, complicating the comparison of study results and limiting generalisability.

Notwithstanding the controversy surrounding whether depression is best classified using a dimensional or categorical diagnostic system, the DSM and ICD classification systems are used globally. Within North America, and the field of psychology, the DSM is more frequently used and will be the system referred to in this study. As such, the manner in which depression is classified according to the DSM is outlined in detail, and diagnostic criteria for a depressive episode used within the ICD-10 will be presented to facilitate comparison between these two prominent classification systems.

DSM-IV-TR

The DSM-IV-TR (American Psychiatric Association, 2000) classifies depressive disorders as being unipolar or bipolar in nature. Unipolar depression comprises a set of signs and symptoms that represent a fixed depressed mood state that does not vary as a consequence of internal or external events (Faravelli, Ravaldi & Truglia, 2005). Unipolar depression is usually episodic, although can be recurrent or chronic (American Psychiatric Association, 2000). Conversely, bipolar depression is characterised by extreme oscillations in mood of variable intensity from depression to hypomania or mania (American Psychiatric Association, 2000). For the purpose of this paper, only unipolar depression is discussed. These are Major Depressive Disorder (MDD), Dysthymic Disorder, and Depressive Disorder Not Otherwise Specified (Depressive Disorder NOS).

The DSM-IV-TR distinguishes Major Depressive from Dysthymic Disorder in terms of duration and severity of symptoms. According to the DSM-IV-TR (American Psychiatric Association, 2000), an episode of Major Depression is diagnosed when five of a listed nine symptoms are present during the same two-week period. Further, these symptoms must represent a departure from previous functioning, and must cause significant subjective distress. One of the five symptoms must be either depressed mood or loss of interest/pleasure (anhedonia). Other symptoms used as diagnostic criteria include appetite and sleep disturbance,

psychomotor agitation or retardation, fatigue and loss of energy, feelings of worthlessness or inappropriate guilt, decreased ability to concentrate, and passive or active suicidal ideation (American Psychiatric Association, 2000). Episodes of Major Depression are not diagnosable disorders in and of themselves, rather they comprise the building blocks for a diagnosis of MDD-Single Episode, or MDD-Recurrent (> 1 episode) (American Psychiatric Association, 2000).

In contrast to MDD, a diagnosis of Dysthymic Disorder requires that an individual experience a chronically depressed mood for most of the day, for more days than not for a period of two years. In these terms, Dysthymia represents a more pervasive depressed mood, and consequently, may be viewed as more typical of the individual's behaviour (American Psychiatric Association, 2000). In addition to a chronically depressed mood, an individual must experience at least two of a possible six symptoms these being: poor appetite or overeating; insomnia or hypersomnia; low energy or fatigue; low self-esteem, poor concentration or difficulty making decisions; and feelings of hopelessness. Like MDD, these symptoms must cause clinically significant distress or impairments in social, occupational or other important areas of functioning (American Psychiatric Association, 2000).

Numerous specifiers for both MDD and Dysthymic Disorder are also presented in the DSM-IV-TR (American Psychiatric Association, 2000). These allow clinicians to specify the current clinical status of the disorder as mild, moderate, severe, in partial remission, in full remission, or chronic. In addition, the condition can be specified through other features of the disorder as, With Catatonic Features, With Melancholic Features, With Atypical Features, or With Postpartum Onset, and may also be described through the course of recurrent episodes as With and Without Full Inter-episode Recovery, or With Seasonal Pattern.

Unipolar depressive disorders contain one other category, Depressive Disorder NOS. This category includes disorders whose essential feature is depressed mood while failing to meet the symptom requirements for Major Depressive or Dysthymic Disorder.

In sum, the underlying assumption inherent within the DSM approach to classifying mental disorders is that psychological disturbance manifests as a syndrome, that is, as a set of signs and symptoms. In these terms, clinical depression comprises a set of cardinal signs (depressed mood, anhedonia) and a suite of associated symptoms (somatic and cognitive disturbances) that may be viewed as consequences of these cardinal signs. In combination, these signs and symptoms comprise the category inclusion criteria for an episode of Major Depression. After meeting the diagnostic threshold (five of nine criteria), depression severity is determined by the total number of criteria met, where the greater the number of symptoms present, the more severe the depression. As such, there is an assumption that a person cannot and does not experience prolonged, intense depressed mood or anhedonia without manifesting a number of cognitive and somatic symptoms. Therefore, irrespective of how qualitatively severe the depressed mood or anhedonic state may be, in isolation, these are not sufficient for a diagnosis of clinical depression.

The Limitations of DSM Depression Classification

Rather than simplifying the identification of individuals experiencing the same psychological problem, the DSM's categorical approach has made the classification system very complex and created numerous problems for clinicians and researchers alike. Some of these problems arise from the reliance on the number, as opposed to the type of symptoms, as the criterion for depression diagnosis.

Specifically, this reliance has led to wide heterogeneity in the experience and presentation of people classified as belonging to the same diagnostic category. This has complicated identification of an absolute threshold for accurate case identification, and may have led to an over-estimation of comorbidity rates since classification is dependent on the co-occurrence of other symptoms. These issues pertaining to the number of symptoms *versus* type of symptoms, the threshold problem, and the artificial inflation of comorbidity rates are outlined in turn.

Number of symptoms *versus* type of symptoms.

As previously described, the DSM-IV-TR regards depression as a syndrome in which an episode of Major Depression comprises a minimum of five symptoms for a two-week period. These symptoms, which constitute the criteria for a Major Depressive episode, are summarised in Table 1 according to the DSM-IV-TR and ICD-10 classificatory systems.

Table 1

Criteria for Major Depressive Episode

	DSM-IV-TR	ICD-10
Symptom Duration	Symptoms nearly every in same 2-week period Change from normal functioning	Episode must last at least 2 weeks, with symptoms nearly every day Change from normal functioning
Core Symptoms	*N*=2 Depressed mood Anhedonia	*N*=3 Depressed mood Anhedonia Fatigue/loss of energy
Auxiliary Symptoms	*N*=7 Weight/appetite loss/gain Insomnia/hypersomnia Observable agitation/retardation Fatigue/loss of energy Feelings of worthlessness/guilt Impaired thinking/concentration Recurrent suicidal thoughts	*N*=7 Weight and appetite change Sleep disturbance Subjective or objective agitation/retardation Low self-esteem/confidence Self-reproach/guilt Impaired thinking/concentration Suicidal thoughts
Inclusion Criteria	One core, 5 symptoms in total PLUS Significant distress OR Social Impairment	Mild episode: 2 core, 4 symptoms in total Moderate: 2 core, 6 symptoms in total Severe: 3 core, 8 symptoms in total
Exclusion Criteria	Not substance related Not mixed episode Not organic Not bereavement Not psychotic	No history (ever) of manic symptoms Not substance related Not organic

Note. Exclusion Criteria refers to the factors that must be ruled out, or excluded, when diagnosing an episode of Major Depression. For example, an episode of Major Depression must *not* be etiologically related to use of a substance (Exclusions =Not substance related). If use of a substance precipitated the depressive episode then according to the DSM-IV-TR, this is more appropriately diagnosed as a Substance Induced Mood Disorder (American Psychiatric Association, 2000).

From Table 1 it is clear that the symptoms listed for an episode of Major Depression are decidedly heterogeneous. As a certain number, rather than a

certain constellation, of symptoms must be present for a DSM diagnosis to be made, the presentation and functional capacity of people can differ so much that some researchers have proposed that people within the same diagnostic class may suffer from different conditions (Becker, Leber & Youll, 1995). Table 2 illustrates this.

Table 2
Possible Clinical Presentation of Two Individuals with DSM-IV-TR Classified Depression

Individual A	Individual B
Depressed Mood	Anhedonia
Difficulty waking	Insomnia
Fatigue	Difficulty making decisions
Weight gain	Agitation
Recurrent suicidal thoughts	Appetite loss

Table 2 highlights the extent to which the experience and presentation of two individuals classified within the one category, depression, may differ. It would appear that the two cases in Table 2 have little in common other than the category inclusion criteria (one core, five symptoms in total). In these terms, symptom heterogeneity and the use of symptom quantity as the threshold criterion for category inclusion may result in a multitude of sub-sets of individuals with dissimilar experiences, distinctly different clinical presentations, and moreover, who appear to be suffering from differing conditions. In these terms, the resulting within class heterogeneity may complicate optimal treatment selection and delivery and possibly affect treatment outcome.

In contrast to the DSM-IV-TR, the ICD-10 classificatory system appears to operate within a mixed framework in which both the number of symptoms and type of symptoms are used as category inclusion criterion and the combination of these indicates depression severity.

The inclusion of severity in the classification of depression may be viewed as an attempt to reconcile the dimensional nature of affective experience with the tradition of viewing mental disorders as categorically distributed. However, inspection of Table 1 reveals that, according to the ICD-10, depression severity depends on the overall symptom total (core + auxiliary), where higher symptom totals indicate greater depression severity. In this way, the continued reliance on the number of symptoms to denote depression severity means that the ICD-10 fails to address affective quality or intensity as meaningful indicators of depression severity, and in so doing, may render the integration of categorical and dimensional classificatory approaches unsuccessful.

The threshold problem.

The DSM classification system has led to the issue of subthreshold disorders. According to Ingram and Siegle (2002), subclinical (subthreshold) depression denotes:

> Any state in which depressive symptoms are present, but in which they are not present in either sufficient number or severity to qualify as clinical depression/Major Depressive Disorder, or in which they have not been assessed in a manner that would allow for a determination of whether a diagnosis of MDD is warranted (e.g. self-report questionnaires) (p. 90).

The DSM classificatory system requires that clinicians use categorical concepts as diagnostic threshold criteria to determine who is sufficiently unwell to justify treatment. In an attempt to address the problem of subthreshold depression, new categories (Dysthymia and Depressive Disorder NOS) were added to the DSM classification of depression and were intended to encompass subthreshold forms of Major Depression. Despite these additions, epidemiological studies have found that more than one third of patients presenting with depressive symptoms that caused significant psychosocial impairment continued to fail to meet DSM diagnostic criteria for any category of depressive disorder (Judd, Akiksal & Paulus, 1997; Judd, Rappaport, Paulus & Brown, 1994). For example, researchers investigating one-month point prevalence rates of depressive symptoms in the

general population (*N*=10,526) found that 39% of respondents who reported depressive symptoms reported subthreshold depression (two or more depressive symptoms below DSM diagnostic threshold) that was associated with harmful dysfunction in five of six measures of adverse outcome. Moreover, they had a significantly increased prevalence of histories of Major Depressive episodes, and had an elevated lifetime prevalence of suicide attempts (Judd, Akiksal & Paulus, 1997).

In line with these results, other researchers have found subthreshold depression to be associated with increased disability (Williams, Kerber, Mulrow, Medina & Aguilar, 1995), to be an antecedent to clinical depression (Akiksal, Judd, Gillin & Lemmi, 1997; Angst & Merikangas, 1997; Judd et al., 1998), and to have a negative impact on the course of comorbid disorders with regard to duration, chronicity (Levenson, Hamer & Rossiter, 1990; Saravay & Lavin, 1994), disability (Maier, Gansicke, Weiffenbach, 1997), work impairment (Wittchen, Nelson & Lachner, 1998), QoL, and suicide (Lecrubier & Üstün, 1998). In this context, it is clear that subthreshold depression has important implications for those people who fail to meet diagnostic criteria, but are nevertheless functionally compromised, and consequently may go untreated.

Inflation of comorbidity rates.

The extent to which changes in the diagnostic systems used in almost all studies of comorbidity has resulted in an over-estimation of comorbidity rates in clinical and community samples remains controversial (Frances et al., 1992). Commencing with the DSM-III (American Psychiatric Association, 1980) and continuing through the DSM-IV-TR (American Psychiatric Association, 2000), the DSM classification system dramatically increased the number of diagnostic categories and reduced the number of exclusion criteria. As a justification, Spitzer, Williams, Gibbon and First (1992) assert that the intention of the changes to retain potentially important differentiating information that could improve understanding of aetiology, course and likely response to treatment. However, the introduction of these changes meant that many people who would have received only a single diagnosis in previous systems now receive multiple diagnoses (Kessler, Gottlib & Hammen, 2002). As such, it could also be argued that these

changes had the unintended negative consequence of artificially inflating the estimated prevalence of comorbidity (First et al., 1990) thereby reducing the accuracy of epidemiological studies, and interfering with optimal health care planning, resource allocation, and complicating appropriate treatment selection.

An Alternative Definition of Depression

There is another way in which depression may be defined. The term depression originates from Old French *depresser* meaning to depress, from Latin *depressum* pressed down, from de- down + premere to press + -ion indicating an action, process or state (Colman, 2003). While the derivation of the word 'depression' may be of historical interest, its linguistic roots remain relevant to the contemporary use of the word in the general community. It is also relevant to the fields of psychology and medicine where 'depression' continues to broadly denote a lowering of mood. This has important definitional implications. In order to determine whether an entity has been lowered or pressed down, one must first establish the normative parameters for that entity.

Within a medical framework, the threshold for depression cannot be used to establish the normative parameters for mood, rather cut-scores on depression inventories are supposed to delineate the presence of disorder/disease from the absence of disorder/disease. However, an absence of disorder does not define the normative parameters for mood, as it cannot infer a normal level of psychological functioning, that is, a state of mental health. The WHO's definition of health is referred to as the gold standard definition of the term and remains unchanged since 1948. Here, health is defined as 'a state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity' (WHO, 1948). According to this definition, three areas of wellbeing combine to represent complete health. If complete health assumes the presence of physical, mental and social wellbeing, then it logically follows that each must encompass the normative parameters for that area. In these terms, the normative parameters for mood may be defined by the presence of wellbeing in this area and should represent a relatively stable state or range within which an individual ordinarily functions. It is proposed that this stable state or range is best described in terms of perceived wellbeing. Further, if normal positive levels of personal wellbeing comprise the

normative parameters for mood, then it is possible that, contextualised by the roots of the term depression described above, depression, representing a lowered state or lowered normal levels of mood, is best described in terms of low levels of personal wellbeing. This conceptualisation of depression will be explored in the second part of this chapter in which the relationship between depression and SWB is explored.

Summary of Issues of Definition

In summary, differences in the definition of depression appear to depend on whether the term is used to describe affective states and mood (general community/psychology) or whether depression is conceptualised as a syndrome comprising negative affect, negative cognitions and an assortment of somatic symptoms (medicine/psychiatry). Nevertheless, despite these differences, the element common to all definitions is the idea of a lowered affective state that represents a departure from a normal level of functioning.

Epidemiology

Epidemiological research into depression is vital for several reasons. First, assessing the magnitude of the burden of disease in the general population allows for effective health care planning and enables optimal allocation of resources. Second, depression has been found to cause poorer outcomes in people with comorbid mental and physical disorders (American Psychiatric Association, 2000) and information regarding the frequency of depression in these populations informs treatment selection and planning. Third, epidemiological research adds to clinical research by highlighting vulnerable populations and illuminating the course of depressive disorders. Finally, information on the prevalence of depressive disorders, including subthreshold forms of depression in the general population, can be used to investigate the degree of congruence with the distribution of SWB.

Measurement

The frequency of depression can be measured in a number of ways. These include incidence, point prevalence, period prevalence, and lifetime prevalence.

Bebbington (2004) defines these terms as follows: incidence refers to the number of cases in a given period as a proportion of a population; point prevalence is the number of cases identified at a point in time as a proportion of a total population; period prevalence rates represent the number of cases identified as in existence during a specified period as a proportion of a total population; and lifetime prevalence is similar to period prevalence where the period of case identification comprises the entire lifetime of each person at the point of ascertainment. While most studies report point and lifetime prevalence rates, twelve-month period prevalence rates are more useful for giving an indication of the number of people who may be seeking treatment in any given year (WHO, 2001), thereby facilitating effective health care planning and resource allocation.

Epidemiological studies of depression in the community, and primary care settings, use various methods of assessment to identify cases. These may be through the use of clinician rated or self-report screening questionnaires, diagnosis by primary care professionals, interview by lay people and clinicians, or less frequently, some combination of these. Further, the questionnaires, clinical interviews (semi-structured *versus* fully structured) and classification systems (ICD *versus* DSM) used for case identification commonly vary between studies. This variability in definition and methodology has led to confusion regarding the true prevalence of depression in the general population, with seemingly minor changes in assessment criteria and methods of assessment producing considerably different results (Regier et al., 1998; Narrow, Rae, Robins & Regier, 2002). However, some large-scale epidemiological studies and reviews that group studies using similar methods have made progress in reliably comparing prevalence rates for depressive disorders between nations (Goldberg & Lecrubier, 1995, Weissmann et al., 1996) and within nations (Weismann et al., 1991).

These latter studies are referred to as third generation studies and represent a major advancement from earlier epidemiological research studies. According to Dohrenwend and Dohrenwend (1982), community residents were not contacted in first generation epidemiological research studies, rather case identification relied on the use of agency records. The major improvement in second generation studies (post-World War II), was that community residents were directly

interviewed by clinicians who gave a diagnosis, or were surveyed using forms that generated global ratings of psychopathology. Third generation, or current, studies aimed to improve the reliability of psychiatric diagnoses by decreasing the potential for disagreement between interviewers/clinicians using standardised diagnostic criteria sets (DSM, ICD) and structured clinical interviews. The prevalence rates from select reviews and third generation epidemiological studies for unipolar depressive disorders in the community and primary care settings studies are presented below.

Prevalence Rates in Community Samples

According to the DSM-IV-TR (American Psychiatric Association, 2000), the lifetime prevalence rates of MDD in adult community samples ranges from 10-25% for women and 5-12% for men. Point prevalence rates are lower for both genders, ranging from 5-9% for women and 2-3% for men (American Psychiatric Association, 2000). For Dysthymic Disorder, the lifetime prevalence rate for men and women combined is approximately 6% with a point prevalence of 3% (American Psychiatric Association, 2000). Importantly, the prevalence rates cited for MDD presented in the DSM-IV-TR are reportedly independent of education, income, marital status and ethnicity (American Psychiatric Association, 2000). In line with many other research findings (Jacobi et al., 2004), the prevalence rates reported in the DSM-IV-TR rates indicate that depression is more common in women than in men (approx. 1.5 to 2.5 times higher for women). However, the prevalence rates for depressive disorders presented in the DSM-IV-TR are not referenced and there is no description of the research design or methodology used to identify cases. Despite these rates being repeatedly cited within much of the depression literature, failure to include information on the sources from which the prevalence rates were derived means that it is not possible to assess their veracity.

Point prevalence rates reported in other large-scale studies differ from those reported within the DSM. For example, in 2000 the WHO undertook a comprehensive assessment of the global burden of disease, GBD2000. Results were based on extensive analysis of mortality data for all regions of the world coupled with systematic reviews of epidemiological studies and population-based mental health surveys. In this study, estimates for point prevalence of unipolar

depressive episodes were 3.2% for women and 1.9% for men, both substantially lower than those reported in the DSM (5-9% for women; 2-3% for men). One-year prevalence rates for women and men were 9.5% and 5.8% on average respectively (Murray & Lopez, 2000).

Lifetime prevalence rates vary widely across nations. Taking only those studies that used fully structured interviews using the Composite International Diagnostic Interview (CIDI) (WHO, 1992) to identify cases of depressive disorders, lifetime prevalence estimates for Major Depression range from 3.4% in the UAE (Abou-Saleh, Ghubash & Daradkeh, 2001), to 16.2 % in the US (Kessler et al., 2003), 16.8% in Sao Paulo, Brazil (Andrade et al., 2002), and up to 17.8% in Norway (Kringlen, Torgersen & Cramer, 2001). In line with these results, Weissmann et al. (1996) summarised depression prevalence from ten population-based studies that used similar case identification methods.

In each of the ten studies summarised by this author, cases of Major Depressive episode were identified using the Diagnostic Interview Schedule (DIS) (Robins et al., 1985) originally developed for use in the US National Institute of Mental Health Epidemiological Catchment Area Program surveys (Robins, Helzer, Croughan & Ratcliff, 1981). The DIS is a fully structured interview that may be conducted by lay interviewers, making this instrument more cost effective than those requiring professional clinician interviewers (i.e. semi-structured interviews). Nevertheless, fully structured interviews, such as the DIS, do not allow for an exploration of reported symptoms. In these terms, depressive symptoms may be recorded as present when subsequent clinical enquiry might reveal otherwise, and *vice versa*.

As few instruments cover the full range of diagnoses, validity estimates for the DIS have been derived through comparison of DIS diagnoses by lay interviewers with those obtained by trained professional clinicians, such as psychiatrists and psychologists. Kappa values of agreement have been disappointing, for example, Anthony et al. (1985) reported a mean of only .15 across diagnoses. Notwithstanding these limitations, the prevalence rates from the ten studies presented by Weissman et al. (1996), may be compared as a means to examine the presence or otherwise of cross-national differences in prevalence

rates. Table 3 contains the one-year and lifetime prevalence of a major depressive episode summarised by this author.

Table 3

One-year and Lifetime Prevalence Rates for Major Depressive Episode from Studies Using the DIS (Weisman et al., 1996)

Site	Author	*N*	One Year	Lifetime
Beirut, Lebanon	Karam, 1992	528	-	19.0
Christchurch, NZ	Oakley-Browne et al., 1989	1 498	5.8	11.6
Edmonton, Canada	Bland et al., 1988	3 258	5.2	9.6
Florence, Italy	Faravelli et al., 1990	1 000	-	12.4
Korea	Lee et al., 1990	5 100	2.3	2.9
Paris, France	Lepine et al., 1989	1 746	4.5	16.4
Puerto Rico	Canino et al., 1987	1 513	3.0	4.3
Taiwan	Hwu et al., 1989	11 004	0.8	1.5
United States	Robins & Regier, 1991	18 571	3.0	5.2
West Germany	Wittchen et al., 1992	481	5.0	9.2

Despite Weismann and her colleagues concluding that the variation between nations is not great, inspection of the prevalence rates in Table 3 indicates that the one-year prevalence rates vary five percentage points (0.8%-5.8%) and the lifetime rates vary 17.5 points from 1.5% in Taiwan (Hwu et al., 1989) to 19.0% in Beirut (Karam, 1992). However, the variation in prevalence rates in Table 3 is not easily explained. The two lowest rates are from Asia. The high rates in Lebanon are perhaps understandable in terms of increased exposure to traumatic events, threats of violence and stressors but the rates in France, Italy, and New Zealand are also high. Further, in contrast to other studies that have found very high rates in South American centres (Andrade et al., 2002), the lifetime prevalence of depression in Puerto Rico is not high (4.3%). With the exception of the particularly low rates found in Taiwan (0.8%; 1.5%) and Korea (2.3%; 2.9%), the overall variation in prevalence rates does not seem to

correspond to any obvious cultural difference between the location of the surveys, although it remains possible that there are ethnic differences in vulnerability to or reporting of depression (Weismann et al., 1996).

In an attempt to explain the particularly low rates of depressive disorders found in Asian countries, epidemiological researchers have proposed cultural response bias as one possible explanation (Weismann et al., 1996; Simon, Goldberg, Von Korff & Üstün, 2002). These researchers argue that low social acceptability for expressing emotions, traditional values of 'saving face' (Iwamasa, Hilliard & Osato, 1998), and linguistic differences in the tendency to report distress (Dohrenwend, 1966) may decrease the likelihood that people from Asian countries will report negative affect and cognitions. Alternatively, or perhaps additionally, the depressive syndrome, as defined in Western Europe and the US may exist in different forms, or may not exist at all, in other cultures, and cross-cultural application of DSM and ICD criteria sets may yield erroneous prevalence rates (Littlewood, 1990; Mezzich et al., 1999). Therefore, rather than indicating a relative absence of depressive disorders, the low prevalence rates found in Asian countries may result from a poor fit between Western definitions and measurements of depression and the expression of depression in eastern cultures. However, whether the overall variability of prevalence rates between nations reflect true differences in rates of depression, are the consequence of cultural differences, or represent problems with depression definition and measurement is not known.

Australian Samples

In Australia, Andrews, Henderson and Hall (2001) investigated the prevalence of psychiatric disorders and associated comorbidity, disability and service utilisation in the general population using a large sample (10,641) of general population adults drawn from a national survey conducted by the Australian Bureau of Statistics (ABS). Psychiatric disorders were identified using an automated presentation of the CIDI (WHO, 1997) .The CIDI is a structured interview that uses both ICD-10 and DSM-IV (American Psychiatric Association, 1994) diagnostic criteria to identify and classify disorders, affording the opportunity to compare prevalence rates generated by differing classification

systems. Andrews et al. reported one- and twelve-month prevalence rates for all mental disorders, and those found for Major Depression and Dysthymia are summarised in Table 4.

Table 4
Twelve-month and One-month Prevalence Rates for Major Depression and Dysthymia According to ICD-10 and DSM-IV Criteria (Andrews et al., 2001)

Disorders	ICD-10				DSM-IV			
	12 month		1 month		12 month		1 month	
	% (s.e.)		% (s.e.)		% (s.e.)		% (s.e.)	
Major Depression	6.7	(0.4)	3.3	(0.2)	6.3	(0.3)	3.2	(0.2)
Dysthymia	1.3	(0.2)	1.1	(0.2)	1.1	(0.1)	0.9	(0.1)

Results summarised in Table 4 demonstrate that prevalence rates produced by the ICD-10 classification were higher (+ 0.1 to +0.4 points) than those produced by DSM-IV criteria. When prevalence rates for Major Depression and Dysthymia are combined, 7.4% of Australians met DSM-IV criteria for a depressive disorder in the last twelve months and 4.1% in the last month.

Interestingly, the one month prevalence rate for unipolar depressive disorders (4.1%) found by Andrews et al. (2001) approximates the proportion of people in the Australian general population who report being dissatisfied with their life (Davern, 2004) scoring less than 50 percentage points on the PWI-A (4.4%) (Cummins et al., 2003). In this context, the proportion of people experiencing depression and life dissatisfaction may be seen as virtually equivalent. This raises the question of whether depression and life dissatisfaction co-occur, or moreover, as is contended in this thesis, whether suppressed levels of normal positive HPMood (PWI-A<50 points) constitutes depression.

Although effective treatments for depression are available throughout Australia, half of the people who met criteria for a mental disorder did not seek help (Andrews et al., 2001). Since publication of this research, public education initiatives (e.g., Beyond Blue 2000) aimed at normalising depression, removing

the stigma associated with mental health issues and encouraging treatment seeking behaviour, particularly in men, have commenced in Australia. Data pertaining to the impact such initiatives may have had in the general community are not yet available.

Prevalence Rates in Primary Health Care Settings

In primary care settings, research on the prevalence of mental disorders conducted by the WHO (Ustin & Sartorius, 1995; Goldberg & Lecrubier, 1995) and reported in the World Report on Mental Health (WHO, 2001) compared prevalence rates across countries. In this research, cases were identified from fifteen primary care centres in fourteen countries using three methods of assessment, a brief self-report screening instrument (General Health Questionnaire, GHQ-12) (Goldberg, 1992), a detailed structured interview (CIDI), and a clinical diagnosis by a primary care physician (PCP). As the rank ordering of centres on the prevalence of mental disorders did not vary as a function of case identification, Goldberg and Lecrubier (1995) concluded that there were true differences in prevalence between the centres. The prevalence rates for all mental disorders by assessment method for each of the fifteen centres are summarised in Table 5. Additionally, the prevalence rates for current depression according to the CIDI are presented in Table 5.

Inspection of observed prevalence across nations in Table 5 reveals a pattern of high rates in South American centres, intermediate rates in European and US centres, and low prevalence rates once again in Asian countries. This pattern is consistent with results from several other community and primary care surveys (Araya, Wynn, Leonard & Lewis, 1994; Araya, Rojas, Fritch, Acuna & Lewis, 2001; Lima, Beria, Tomasi, Conceicao & Mari, 1996; Weismann et al., 1996; Almeida-Filho et al., 1997).

From Table 5, it is clear that the rank ordering of centres on prevalence of all mental disorders is similar across methods of identification suggesting that, in line with Goldberg and Lecrubier's (1995) interpretation, there are true differences in prevalence between the centres. However, the prevalence rates for

all mental disorders within centres differ substantially depending on the method of case identification used (CIDI *versus* PCP diagnosed *versus* self-report inventory).

Table 5

Prevalence of all Mental Disorders Identified by CIDI, PCP, GHQ Assessment and Current Depression by CIDI (Goldber & Lecrubier, 1995)

Country	All mental disorders				Current Depression	
	CIDI	PCP	Self-Report	Rank	CIDI	Rank
	(%)	(%)	(%)		(%)	
Santiago, Chile	52.5	58.5	51.1	1	29.5	1
Rio de Janiero, Brazil	35.5	20.4	20.9	2	15.8	4
Paris, France	26.3	30.3	29.5	3	13.7	5
Manchester, UK	24.8	28.1	27.5	4	16.9	2
Groningern, Netherlands	23.9	27.0	28.2	5	15.9	3
Mainz, Germany	23.6	33.0	30.3	6	11.2	7
Bangalore, India	22.4	16.2	23.9	7	9.1	8
Athens, Greece	19.2	11.6	16.1	8	6.4	9
Berlin, Germany	18.3	32.5	28.4	9	6.1	11
Ankara, Turkey	16.4	8.2	12.2	10	11.6	6
Seattle, US	11.9	19.7	19.5	11	6.3	10
Verona, Italy	9.8	47.3	23.3	12	4.7	12
Ibadan, Nigeria	9.5	27.9	15.7	13	4.2	13
Nagasaki, Japan	9.4	4.8	13.3	14	2.6	15
Shanghai, China	7.3	5.4	13.9	15	4.0	14

There is growing acceptance of the CIDI as the benchmark method for the identification of psychiatric disorders among epidemiological researchers, as cases may be identified using either ICD or DSM diagnostic criteria in a fully structured, standardised interview. As such, the prevalence rates generated by the CIDI have been deemed more reliable for identifying mental disorders than diagnosis by a PCP or self-report screening inventory. It is of interest to note that, of the three methods used, the CIDI produced the highest prevalence rates in three

centres (Rio, Ankara and Athens) the lowest rates in nine centres (Paris, Manchester, Groningern, Mainz, Bangalore, Berlin, Seattle, Verona and Ibadan) and rates that fell between those produced by PCP diagnosis and self-report inventories in three centres (Santiago, Nagasaki and Shanghai).

If the rates produced by the CIDI reflect the truest indication of depression prevalence, then PCP and self-report inventories must be viewed as either over- or under-estimating the prevalence of mental disorders by centre. For example, in Japan and China, relative to the prevalence rates derived from the CIDI, PCP under-estimate (-4.6%; -1.9%) and self-report inventories over-estimate (+3.9%; +6.6%) the prevalence of mental disorders, whereas the reverse is true in Chile (PCP +6.0%; S-Report, -1.4%). In Western European centres (Manchester, Groningern, Mainz, Berlin and Verona) and the US, both PCP and self-report inventories over-estimate the prevalence of mental disorders, with the greatest percentage point discrepancy found in Verona, Italy, where there is a 37.5 point difference between the prevalence rates according to the CIDI (9.8%) *versus* those reported from PCP diagnoses (47.3%).These analyses also revealed that point prevalence rates for current depression varied across countries, from a low of 2.6% in Japan to high of 29.5% in Chile. While it is tempting to infer that the discrepancy between the highest and lowest rates for depression may be accounted for by differences between the countries, due to such influences as development classification (income, economic status, and human assets such as nutrition, health, education and literacy), Goldberg and Lecrubier (1995) found no consistent differences in prevalence rates between developed and developing nations. Inspection of the ranking order for countries in Table 5 verifies this. Three (England, Netherlands, France) out of the five countries with the highest rates of depression are classified as developed countries according to the United Nations (UN) (2008).

In an attempt to explain the differences in observed prevalence between centres, Goldberg and Lecrubier (1995) offer four possibilities. These are that the observed differences reflect true differences in prevalence, differences in concepts of illness, differences in tendency to seek help in each culture, or demographic differences in the help-seeking populations (Golberg & Lecrubier, 1995).

While all of these possibilities deserve serious consideration, the WHO's study (Goldberg & Lecrubier, 1995) is not without serious methodological limitations. First, the three centres with the highest rates of depression (Santiago, Rio, and Manchester) were public clinics serving socio-economically deprived areas. In these terms, the observed differences in prevalence may reflect the specific primary care facilities studied rather than regions or national populations. Second, interviewers were drawn from the same linguistic and cultural groups as the patients and, despite the use of a fully structured diagnostic interview, experimenter, in this case, interviewer effects must impact the WHO's data to some degree. Finally, Goldberg and Lecrubier did not discuss whether the consequences of reporting depression, be they negative (e.g., social stigma or ridicule), or positive (e.g., social or financial support) varied between centres, and therefore may have impacted patients' tendency to report depressive symptoms. While these limitations may constrain the extent to which these data can be interpreted as evidencing true differences in observed prevalence, it may be interesting to speculate, from the perspective of SWB homeostasis, as to why the rates for all mental disorders and current depression are so high in Chile.

Prevalence rates in Chile.

Chilean prevalence rates reported by Weismann et al. (1996) for all mental disorders and current depression according to the CIDI are a staggering 52.5% and 29.5% respectively. These rates are substantially higher (+20.0 points; +12.6 points) than those found in any of the other nine centres.

Over the last three decades, Chile has undergone massive social, economic and political change, as well as significant health reform. These changes have been so great that Chile has been likened to a social laboratory (Unger, De Paepe, Cantuarias & Herrera, 2008) having experienced Christian democratic reformism (1964-1970), democratic socialism (1970-1973), neoliberal authoritarianism (1973-1989), and three democratic coalition governments from 1990 to the present (Silva, 2005).While a return to democracy over the last decade has meant that expenditure in the public health sector has increased (Annick 2002), Chile continues to share the dubious distinction of having one of the most regressive patterns of income distribution in the world (Unger et al., 2008). Despite a

reduction in the percentage of people living below the poverty line, from 45% in 1985 to 21.7% in 1998, income inequality rose, as did inequities in access to health care, decreasing cohesion between the rich and poor, young and old, sick and healthy (Unger et al., 2008).

In the context described above, it is likely that the Chilean people recruited from a public health service and surveyed in 1995, would have had less resources, relative to those with private health cover, to deal with rapid change and the consequent life stressors and challenges. In these terms, it is possible that in the context of escalating social, financial, and health care inequalities, the Chilean people surveyed had not been able to adapt adequately to changing life circumstances. Failure to adapt means that in an intensely challenging context, the systems that ordinarily function to maintain normal levels of mental and physical health may be overwhelmed and fail. In terms of mood, if normal levels of mood are represented by normal positive levels of SWB as proposed within this thesis, then in line with the theory of SWB homeostasis (Cummins et al., 2009, in press), the homeostatic defence system that maintains SWB levels within individual set-point ranges will be challenged. If the challenging circumstances are of sufficient strength, and are sustained, then the homeostatic system may be overwhelmed and fail. The consequence of this is the suppression of the normal positive sense of HPMood, and the subsequent loss of SWB.

While this appears a reasonable explanation for the extraordinarily high rates of mental disorders and depression reported in Chile, whether the rate and magnitude of social and economic change has been great enough, relative to other countries, to account for such high prevalence rates, requires determination.

Taken together, the difficulty in standardising the methodology used in epidemiological research studies and problems related to the definition and measurement of depressive disorders across cultures continues to hinder understanding. These problems fundamentally limit researchers' ability to accurately assess and describe the prevalence of depressive disorders in the general population and primary care settings. In these terms, prevalence rates for depression reported in epidemiological studies to date must be viewed as estimations, reflecting only approximations of the true incidence of depression in

the general population. As such, projections derived from epidemiological data and health economic measures, to describe the impact and economic cost of depression, must be viewed as rough estimates.

The Impact of Depression

A collaboration between the Harvard School of Public Health, the World Bank and WHO introduced new measurements, the Disability Adjusted Life Years (DALYs) and Years Lived with Disability (YLDs) in an attempt to quantify the burden of disease (Murray & Lopez, 1996). DALYs represent the sum of years of potential life lost due to premature mortality and the years of productive life lost due to disability. One DALY may be thought of as one year of healthy life lost, and the burden of disease as a measurement of the gap between an ideal situation in which everyone lives into old age free of disability and disease and current health status according to objective indictors (WHO, 2001). In 2000, the WHO conducted the largest, most comprehensive review of mortality and morbidity by age, gender and region, known as of the Global Burden of Disease (GDB) 2000 (Murray, Lopez, Mathers & Stein, 2001). Using a structured diagnostic interview to detect cases of depression (CIDI) (WHO, 1997), this research group found depression to be the leading cause of YLDs, accounting for 11.9% of total YLDs and is the fourth leading contributor to the global burden of disease, accounting for 4.4% of the total DALYs (Murray & Lopez, 2000). In the age category fifteen to 44 years for both sexes combined, depression accounts for 8.6% of DALYs lost, making it the second leading cause of disability of all diseases (Murray & Lopez, 2000; WHO, 2001). Projections based on epidemiological data estimate that by the year 2020, depression will account for 5.7% of the total burden of disease, second to ischemic heart disease as the leading cause of DALYs lost (WHO, 2001). Of course, all these conclusions are at the mercy of the uncertainty concerning depression measurement, as has been described.

Economic Costs

In the US, MDD is already the leading cause of disability, surpassing chronic medical conditions such as hypertension, heart disease and diabetes in this

respect (Druss, Rosenheck & Sledge, 2000). The economic burden of this disease in the US alone was $83.1 billion in 2000: $26.1 billion in direct medical costs; $5.4 billion in suicide-related mortality costs; and $51.5 billion in workplace losses (Greenberg et al., 2003).

In Australia, per capita costs for the burden of depression in South Australia are comparable to those found in the US, with a total financial cost of $2.8 billion per annum (Hawthorne, Cheok, Goldney & Fisher, 2003).

Summary of Epidemiology

In summary, despite the development of standardised diagnostic criteria sets and assessment methods in epidemiological research, problems with depression definition, measurement and the application of assessment methods across cultures is limiting researchers' abilities to describe the frequency and magnitude of depressive disorders world-wide. Nevertheless, based on extant epidemiological data, projections of the global burden of depressive disorders highlight that depression is currently one of the most pervasive and pressing health concerns world-wide. In these terms, it is clear that the manner in which depression is currently defined and classified requires urgent review. Additionally, in order to facilitate the prevention, early detection and effective treatment of depression, it is vital that depression instruments are psychometrically sound and that the manner in which these instruments operationalise to measure depression has validity. The measurement of depression will now be reviewed.

Depression Measurement

The items that are used to measure depression may be broadly divided into two categories: those that assess diagnostic criteria and those that assess symptom severity. Most instruments combine both.

Instrument selection may depend on context and purpose, and thus, instruments may be further categorised into clinician-rated *versus* self-rated measures. Generally, clinical interviews are conducted by specially trained mental health professionals (e.g., psychiatrists and psychologists) for diagnosing and

treating depression as a disorder. Self-report measures are more commonly used to assess symptom severity, to guide and evaluate treatment and to assist in case identification in primary care and research settings.

In this section, a brief overview of both diagnostic interviews and self-report questionnaires is provided and the empirical literature related to their reliability and validity is reviewed. The three self-report instruments used in this study are then sequentially presented and evaluated. These are, the BDI-II (Beck, Steer & Brown, 1996), the depression scale of the DASS (DASS-D) (Lovibond & Lovibond, 1995), and the depression scale of the HADS (HADS-D) (Zigmond & Snaith, 1983).

Diagnostic Interviews

Structured and semi-structured interviews are designed to standardise and improve the reliability of psychiatric diagnoses (Dozois & Dobson, 2002) through manualised assessment of DSM and/or ICD diagnostic criteria. Two of the most widely used diagnostic interviews that encompass all categories of mental disorders, are the WHO's fully structured CIDI (WHO, 1992), and the semi-structured Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) (First, Spitzer, Gibbon & Williams, 1992). Both the CIDI and the SCID-I afford clinicians the opportunity to specify distinguishing characteristics (e.g., MDD *with melancholic features*), in addition to the severity of depression (mild, moderate or severe). Nevertheless, their primary purpose is to detect the presence or absence of symptoms listed in criteria sets within in the DSM or ICD.

Comparison of the CIDI and SCID-I.

The CIDI was developed by the WHO to assist psychiatric epidemiologists to carry out community surveys. It is a fully structured interview, and as such, may be conducted by a specially trained lay-interviewer. It does not allow any deviations from the formatted questions, and does not allow the interviewer to restate, rephrase, probe for clarification or challenge respondents. In these terms, symptoms may be recorded as present when subsequent clinical enquiry might reveal otherwise and *vice versa*, leading to a potential for over- or under-diagnosis of mental disorders respectively.

In contrast to the CIDI, the semi-structured SCID-I provides a standardised clinical interview to determine DSM-IV (American Psychiatric Association, 1994) diagnoses and differential diagnoses. While the SCID-I is structured to match DSM-IV criteria, unlike the CIDI, it also utilises the skills of professional trained clinicians (psychiatrists or psychologists) to explore responses, clarify items, repeat, rephrase and follow-up responses in order to fully understand the nature and extent of the reported symptoms (First et al., 1992). In this context, the SCID-I requires that the interviewer have a higher level of clinical expertise to ensure the reliable administration of the instrument.

While the SCID-I uses the DSM classification system, the CIDI allows the interviewer to diagnose disorders according to either, or both, DSM or ICD diagnostic criteria. Both interview schedules have undergone numerous revisions since their original development, and both interviews are now available in various administration formats. The SCID-I is available in a computer-assisted format to aid the interviewer, while the CIDI is available in a completely automated format and an extended version that takes up to two hours to complete (the World Mental Health CIDI or WMH-CIDI) (Kessler & Üstün, 2004).

The time taken to complete these diagnostic interviews may vary depending on the version used, and the extent to which SCID-I items require follow-up. In general, full administration of the SCID-I-I takes 45-90 minutes, whereas the CIDI may take 45-120 minutes for full administration (Dozois & Dobson, 2004). Clearly, the time and resources required to administer these interviews fully may be unattainable in many research settings, and may mean that these interviews are not viable options as screening tools in primary care settings.

Notwithstanding these administration difficulties, both the CIDI and SCID-I contain discrete modules of formatted questions, with each module assessing a particular category of mental disorder. That is, one module assesses mood disorders, another assesses anxiety disorders and so on. As such, it is not necessary for clinicians or researchers to administer the entire interview schedule in each instance. For example, a clinician working in a depression and anxiety disorder programme, might only be interested in assessing the client for

programme eligibility, and therefore administer only those modules that assess mood disorders and anxiety disorders. While this approach may increase the utility of the instrument by lowering administration costs in terms of time and resources required to complete the interview, it may also lead to a high rate of false-positive cases (Kurdyak & Gnam, 2005). This is due to the overlap in diagnostic criteria between disorders from differing categories (Farvelli et al., 2005).

Nevertheless, for the purpose of psychiatric research, the CIDI has some distinct advantages relative to the semi-structured SCID-I. These include an ability to compare differences between the prevalence rates generated by the two established taxonomies for classification of mental disorders (ICD *versus* DSM). This comparison allows for further refinement and evolution of these systems (Golberg & Lecrubier, 1995; Kessler et al., 2004; Kessler & Üstün, 2004). A second advantage of the CIDI is higher levels of inter-rater reliability, due to the fully structured format (Andrews, Henderson & Hall, 2001; Goldberg & Lecrubier, 1995), and lower administration costs through the use of trained lay-person interviewers as opposed to specialist clinician interviewers. As a result of these advantages, the CIDI is increasingly referred to as the 'gold standard' for estimating prevalence rates internationally (Andrews et al., 2001; Komiti et al., 2001, Goldberg & Lecrubier, 1995; Üstün & Sartorius, 1995).

Conversely, critics of the CIDI argue that the instrument's reliability is yet to be conclusively demonstrated (Kurdyak & Gnam, 2005) and caution against the adoption of this instrument as a diagnostic 'gold standard' (Andrews & Peters, 1998). In a recent review of the CIDI's diagnostic utility, Kurdyak and Gnam (2005) evaluated the evidence supporting the validity of the CIDI, focussing on the extent to which the depression diagnoses reflected true cases of depression. Using data drawn from a nationally representative community survey, the Canadian Community Health Survey: Mental Health and Wellbeing, these authors found that, as expected, the CIDI performed well in terms of inter-rater reliability. However, the CIDI had never been validated in general population samples, rather, its validity had been tested using only clinical samples. As clinical samples have higher prevalence rates of mental disorders relative to community samples

(Faravelli et al., 2005), Kurdyak and Gnam argue that it is likely that the CIDI will generate a high rate of false-positive cases, resulting in the artificial inflation of prevalence rates in the community.

Other problems detailed by Kurdyak and Gnam (2005) relate to the continual modification and refinement of CIDI items, and seemingly idiosyncratic use of differing versions of the instrument. This potentially limits the generalisability of results across studies and raises concerns regarding the reliability of the instrument. Further, the reliability of the CIDI depression module remains contentious. Reliability studies for clinical interview schedules are generally conducted by comparing the agreement among interviewers' ratings and calculating the Kappa Measure of Agreement (*k*). According to Peat (2001, p.228), a kappa value of less than .5 represents poor agreement, .5-.7 represents moderate agreement, .7-.9 represents good agreement, and above .9 represents very good agreement.

The vast majority of the reliability studies on the CIDI have been conducted using clinical samples, yet the CIDI was developed specifically to be used in epidemiological studies of mental disorders in the general population. In a review of CIDI field trials conducted by the WHO (N=575), Wittchen (1994) concluded that the CIDI demonstrated moderate to very good reliability, with kappa coefficients ranging from .67 to .98. For the mood disorders module, inter-rater kappa values were reported as .95 for any depressive disorder, .96 for Dysthymia, and .97 for Major Depression, single episode. However, two independent studies that compared depression diagnoses generated by the CIDI with depression diagnoses generated by specialist clinicians using a semi-structured interview (SCAN) (Wing et al., 1990) found substantially lower reliability and validity (Andrews, Peters, Guzman & Bird, 1995; Brugha, Jenkins, Taub, Meltzer & Bebbington, 2001). Specifically, Andrews et al. found that the agreement between the CIDI diagnoses and those generated by specialist clinicians was very poor, with a kappa of .39 for current depression diagnoses, and Brugha and colleagues found that the optimal performance of the CIDI to detect cases of depression corresponded to a sensitivity of only 50% and

specificity of 90% (Brugha et al., 2001). These latter findings are in stark contrast to the findings from the field trials of the CIDI.

In combination, these findings suggest that issues pertaining to the reliability and validity of the CIDI in community settings are yet to be definitively resolved. Therefore, researchers must continue to demonstrate caution when interpreting data generated from studies that have used the CIDI as a diagnostic tool, and particularly when used to diagnose depressive disorders in general population samples.

The SCID-I for DSM-IV disorders is a relatively new instrument, and consequently, very few studies have investigated the psychometric properties of this version of the interview. As such, the reliability of the SCID-I for DSM-IV disorders is generally inferred from findings related to earlier versions of the interview (e.g., SCID-I for DSM-III-R). Inferring reliability in this way is potentially problematic because the SCID-I undergoes modifications with each successive version, and therefore, the reliability data reported from one version does not necessarily generalise to another version. Nevertheless, inter-rater reliabilities for the Structured Clinical Interview for DSM Disorders (SCID) for DSM-III-R depressive disorders have been found to range from moderate to very good (k = .72 to k = .92) for general mood disorders, and for Major Depression (k = .64 to k = .93) (Segal, Hersen & Van Hasselt, 1994; Spitzer et al., 1992). However, in non-patient samples Williams et al. (1992) found lower inter-rater reliabilities; k = .64 for Major Depression, and k = .40 for Dysthymia.

Like the CIDI, the validity of the SCID depends on the validity of the nosological system upon which it is based. However, extant validity studies for the SCID-I have investigated validity through the concordance between SCID diagnoses and those derived by specialist clinicians not using the interview schedule. Based on studies that have investigated the SCID-I for DSM-III-R disorders using this method, the kappa values of agreement have been found to be lower than those for inter-rater agreement (Dunner & Tay, 1993; Steiner, Tebbs, Sledge & Walker, 1995).

Conclusions

The CIDI and SCID-I are time intensive and resource intensive methods for diagnosing mental disorders in epidemiological research and clinical settings respectively. The resources required to administer these interview schedules may be justifiable, however, the time and cost of administering these interviews may prove prohibitive for many researchers, and in primary care settings. Additionally, while these interviews aim to improve the reliability of diagnoses through the use of standardised criteria sets and structured interview formats, the reliability of the current versions of the SCID-I and CIDI are yet to be clearly established in general population samples. Further, the validity of diagnostic interviews largely depends on the validity of DSM and ICD criteria sets, and ultimately the medical conceptualisation of depression as a psychobiological construct within a categorical framework. If the underlying structure of most, if not all, mental disorders is dimensional as argued in this thesis and by others (Eyesenck et al., 1983) and as found previously (Ruscio & Ruscio, 2002), then the continued reliance on structured diagnostic interviews to measure psychological disturbance appears imprudent. In these terms, it would appear critical that careful consideration be given to the extent to which any single instrument can be utilised as a diagnostic 'gold standard' in all contexts, for all disorders, and across all cultures. This review will now turn to the examination of an alternate method for measuring depression; self-report instruments.

Self-report Measures

The self-report instrument is a method of assessment whereby the participant answers specific questions in the form of a questionnaire or inventory. Typically, this method of assessment is conducted using a paper-and-pencil format, although computerised formats may also be used. Depending upon the questionnaire, this method can be relatively brief, taking as few as five minutes, and may be completed by the client outside of session time.

Originally developed to quantify the severity of depression in diagnosed patients (Beck ,Ward, Mendelson, Mock & Erbaugh, 1961; Zung, 1965), self-report inventories of depressive symptoms are also frequently used as case-finding

tools in clinical and research contexts as they are easy to use and provide an inexpensive method of screening large numbers of people in a time efficient manner. Presently, self-report measures of depression severity are being replaced, as case-finding tools, by self-administered screening instruments that provide a documented DSM-IV diagnosis (Rogers, Adler, Bungay & Wilson, 2005). However, whether self-administered screeners, such as the Inventory to Diagnose Depression (IDD) (Zimmerman & Coryell, 1987), and Primary Care Evaluation of Mental Disorders-Patient Health Questionnaire (PHQ/PHQ-9) (Spitzer, Kroenke & Williams; 1999) can be used reliably measure symptom severity is yet unclear (Martin & Barlow, 2005). While controversial, in both clinical practice and research studies, self-report severity measures continue to be used as case-finding tools. Additionally, self-report measures of symptom severity are frequently used to establish pre-treatment symptom severity baselines, to track symptom severity over time, evaluate treatment effectiveness and are used by clients for periodic self-monitoring in some empirically validated treatments (Beck et al., 1996).

In this section, the clinical utility and research applicability of self-report inventories is outlined in terms of instrument selection, the relative advantages of self-report *versus* clinician rated-measures, and the limitations of the results generated by many of these instruments. Following this, the three self-report inventories germane to the studies within this thesis are presented and critically reviewed. These are, the BDI-II (Beck et al., 1996), the DASS21-D (Lovibond & Lovibond, 1995), and the HADS-D (Zigmond & Snaith, 1983).

The utility and clinical applicability of self-report measures.

The most comprehensive review and summary of measures of depression, and depressive symptomatology is offered in Nezu, Ronan, Meadows and McClure's (2000) *Practitioner's Guide to Empirically Based Measures of Depression*. In this guide, 52 instruments are described and reviewed in terms of their purpose, administration, scoring, interpretation, clinical utility and research applicability. Of these instruments, seventeen are self-report inventories that measure the severity of depression in adults. The clinical utility and research applicability of these instruments were assigned ratings of either 'high' or 'limited' by Nezu et al. For clinical utility, 'high' ratings were given if the

instrument was frequently used in clinical settings. A rating of ‘limited’ was assigned if the user fee was prohibitive, if the instrument was not frequently used in clinical settings, or if the time required for completion of the questionnaire was impractical. In terms of research applicability, instruments were assigned a rating of ‘high’ if they had been used in research studies and were found to be meaningful, and were rated as being of ‘limited’ value if there was an absence of empirical data supporting the instrument. Of the seventeen self-report inventories, sixteen were given a rating of ‘high’ for both clinical utility and research applicability and these instruments are summarised in Table 6.

Inspection of the instruments listed in Table 6 reveals great variability in number of items (nineteen to 118), response formats (forced choice *versus* Likert-type scales), and average time taken to complete the questionnaires (three to 45 minutes). It would appear that some of the instruments in Table 6 do not meet the inclusion criteria for high clinical and research utility specified by Nezu et al. For example, the Inventory for Depressive Symptomatology (Rush, Gullion, Basco, Jarrett & Trivedi, 1996) may take up to 45 minutes to complete, making this instrument impractical in time limited clinical settings. As such, a rating of ‘limited’ may be a more accurate rating of this instrument’s clinical utility. Further, it is unlikely that people with severe depressive symptoms would be sufficiently motivated to undertake and/or complete lengthy questionnaires. Consequently, the inclusion of instruments containing a seemingly excessive number of items, such as the 118-item Multi-Score Depression Inventory for Adolescents and Adults, also appears questionable. Further, Nezu et al. (2000) rated instruments as having high utility if they were used frequently and meaningfully in research and clinical practice, but fail to describe the criteria used to define these terms. Taken together, failure to adequately define the criteria used to rate clinical utility and research applicability, and the inclusion of instruments that appear to contravene the inclusion criteria that are outlined, limits the usefulness of this summary as a guide to aide researchers’ and clinicians’ selection of instruments.

Table 6

Self-Report Instruments Assessing Depressive Symptomatology in Adults with High Clinical and High Research Utility (Nezu et al., 2000)

Instrument	No. of Items	Response format	Time to complete (mins)	Reference
Beck Depression Inventory-II (BDI-II)	21	4-point	5-10	Beck, Steer & Brown (1996)
Carroll Rating Scale for Depression-Revised	61	yes/no	5-10	Carroll (1998)
Depression Anxiety Stress Scales (DASS)	42	4-point	10-20	Lovibond & Lovibond (1995)
Depression Questionnaire	24	yes/no	20	Sanavio, Bertolotti, Michielin, Vidotto & Zotti (1988)
Hamilton Depression Inventory	23	3- and 4-point	10-15	Kobak & Reynolds (1999)
Hopelessness Depression Symptom Questionnaire	32	4-point	10	Metalsky & Joiner (1997)
Inventory of Depression Symptomatology	30	4-point	30-45	Rush, Gullion, Basco, Jarrett & Trivedi (1996)
IPAT Depression Scale	36	3-point	10-20	Krug & Laughlin (1976)
Multi-Score Depression Inventory for Adolescents and Adults	118	true/false	20-25	Berndt (1986)
Positive and Negative Affect Scales	20	5-point	5	Watson, Clark & Tellegen (1988)
Primary Care Evaluation of Mental Disorders	26	yes/no	5-10	Spitzer et al. (1993)
Profile of Mood States	65	5-point	3-5	McNair, Lorr & Droppleman (1992)
Revised HRSD Self-Report Problem Inventory	76	true/false	10-20	Warren (1994)
Reynolds Depression Screening Inventory	19	Varies	5-10	Reynold & Kobak (1998)
State-Trait Depression Adjective Checklist	34	yes/no	3	Lubin (1994)
Zung Self-Rating Depression Scale	20	4-point	5	Zung (1965)

In the absence of any generally accepted guide for the selection of instruments, the decision between using a clinician-rated measure *versus* self-report inventory may be depend on context, purpose and available resources (Dozois & Dobson, 2004).

In order to make this decision, an understanding of the relative advantages and limitations of the various measurement methods would seem necessary. For self-report inventories, the general advantages and limitations of using this method of assessment to measure depression are described below.

Advantages of self-report measures.

Self-report inventories have several advantages over clinical interviews. First, given that one in seven Australian adults meets criteria for a current mental disorder (Andrews et al., 2001), self-report inventories represent a time efficient, cost-effective way to identify those people most at risk of depressive disorders, those who require referral for further assessment via a more resource-intensive structured clinical interview, and those requiring urgent treatment.

Second, one in three people (Greenberg, Stiglin, Finkelstein & Berndt, 1993) do not seek formal treatment for depression. However, it is possible that some of these people do present to primary care settings with somatic complaints. This would appear reasonable given the high co-occurrence of depression and physical disease (American Psychiatric Association, 2000). In such instances, self-report instruments that assess the psychological dimension of depression (e.g., the DASS or the HADS) may be used by physicians not trained to administer a diagnostic interview, or when working within time constraints, to distinguish physical illness from mood disorders, identify possible cases of depression, and aide referral for further assessment and treatment.

In contrast to fully structured diagnostic interviews, self-report inventories may facilitate optimal treatment selection and implementation by specifically measuring symptom severity. This allows specific symptoms to be prioritised and targeted depending on reported severity.

Finally, in large-scale research studies, where lengthy clinical interviews may not be practical or cost-effective, self-report inventories have high utility with very brief instruments performing as well as longer ones (Shade, Jones & Wittlin, 1998).

Limitations of self-report measures.

However, there are limitations in the extent to which self-report inventories may be used to measure depression. As the specificity and sensitivity of instruments varies widely (sensitivity 50%-97%; specificity 51%-98%) (Williams, Pignone, Ramirez & Stellato, 2002) the empirical evidence related to individual depression inventories must be evaluated so that the instrument with the optimal fit to the context and purpose of assessment may be selected.

More generally, the way in which depression is defined has led to the construction of depression inventories that contain fundamental methodological weaknesses. Specifically, the repetition of negative items in self-report depression inventories may have deleterious effects on the affective tone of respondents as they respond to items. This has the potential to result in the artificial inflation of levels of self-reported symptom severity. This proposition will be tested in the second study of this thesis. Additionally, the use of only negative items perpetuates a disease-focussed approach, limiting the interpretability of results. That is, a low score on a depression inventory is simply indicative of a lack of depressive symptoms, and cannot, by default, infer a state of mental health nor inform on an individual's current level of wellbeing.

Further, since the content of the items in many self-report inventories reflects the current medical definition of depression, and its diagnosis as a clinical syndrome, both psychological and physiological constructs are assessed. Item scores are then summed to yield a total depression scale score. Cut-scores are then used to classify scale total scores into categories of depression severity. However, physiological symptoms may vary because of physical health status and/or culture. As such, it would appear that summing scores on physiological items, which could be the consequence of physical ill health as opposed to depression, together with psychological items is methodologically unsound.

Summary of Self-report Instruments

In summary, self-report measures of depression severity are used in both clinical and research contexts to measure change over time, evaluate the effectiveness of clinical treatments, for periodic client self-monitoring, and more controversially, to identify cases in large-scale research studies and clinical settings. User's guides, developed to aid the selection of instruments, consist largely of subjective, albeit expert opinion, on the strengths and weaknesses of differing instruments. Consequently, the selection of instruments must be informed by the context, purpose, sensitivity, empirical evidence relating to the sensitivity and specificity of the instrument, and the resources available for assessment. Additionally, an understanding of the advantages of using self-report as opposed to clinician-rated measures and the methodological limitations of summing disparate constructs within depression scales is critical to the accurate interpretation of results.

Notwithstanding utility, the self-report instruments that are most widely used by clinicians and researchers alike tend to conceptualise depression as a syndrome consonant with criteria listed in the DSM for depressive disorders. In this context, it is not surprising that the BDI-II (Beck et al., 1996), developed specifically to be consistent with the DSM-IV criteria, has become the most widely used self-report instrument for measuring depressive symptom severity in research and clinical settings (Nezu et al., 2000). Other instruments, such as the DASS21-D (Lovibond & Lovibond, 1995) and the HADS-D (Zigmond & Snaith, 1983), use items that assess only psychological constructs. In this way, these latter two instruments operationalise depression in a manner that more closely reflects the essence of the condition, (i.e. a lack of positive affect) (Watson, Clark & Tellegen, 1988). These three self-report inventories, the BDI-II, the DASS21-D, and the HADS-D are used to measure depression in the same cross-sectional sample in Study 2 in this thesis. Commencing with the BDI-II, these inventories will now be critically reviewed.

Beck Depression Inventory-Second Edition (BDI-II)

Over the last 30 years, the BDI (Beck, Ward, Mendelson, Mock & Erbaugh, 1961; Beck, Rush, Shaw & Emery, 1979; Beck et al., 1996) has become the most widely used psychometric measure of depression severity and the standard by which other self-report depression measures are evaluated (Ritterband & Spielberger, 1996). The medical definition of depression and the criteria for its diagnosis as a clinical syndrome are reflected in the content of the items in the BDI. The latest and third version of the instrument, the BDI-II (Beck et al., 1996), represents a revision of the modified BDI-IA (Beck et al., 1979), which replaced the original BDI (Beck et al., 1961). Items were added in the BDI-II (agitation, concentration difficulty and loss of energy), others were removed (body image, work difficulty, weight loss and somatic preoccupation), and substantial rewording of many items was undertaken, so that the instrument more closely corresponded to depressive symptoms listed as criteria for an episode of Major Depression in the DSM-IV (American Psychiatric Association, 1994) (for DSM-IV-TR criteria see Appendix 1).

The BDI-II is used to assess the severity of depressive symptoms in clinical populations and is commonly used to detect the presence of depression in general population samples (Archer, Maruish, Imhof & Pitrowski, 1991; Beck et al., 1996; Camara, Nathan & Puente, 2000; Steer, Beck & Garrison, 1986). However, the BDI-II was not intended to be used as an instrument for the purpose of clinical diagnosis, as according to the authors, the symptoms of depression may form part of other primary diagnostic disorders and discerning this requires clinical judgement (Beck et al., 1996).

BDI-II Constructs, Response Scale and Scoring

The BDI-II is a self-administered inventory that consists of 21 items presented in a multiple-choice format. The inventory takes approximately five to ten minutes to complete and respondents require a fifth- to sixth-grade reading level to adequately comprehend the items (Beck et al., 1996).The 21 items were designed to reflect the cognitive, affective and physiological symptoms, and

performance decrements of depression associated with depression as a mental disorder and refer to the following areas:

1. Sadness
2. Pessimism
3. Past Failure
4. Loss of Pleasure
5. Guilty Feelings
6. Punishment Feelings
7. Self-Dislike
8. Self-Criticalness
9. Suicidal thoughts or wishes
10. Crying
11. Agitation
12. Loss of Interest
13. Indecisiveness
14. Worthlessness
15. Loss of energy
16. Changes in sleeping patterns
17. Irritability
18. Changes in appetite
19. Concentration difficulty
20. Tiredness or fatigue
21. Loss of interest in sex

Respondents are asked to rate the intensity of the symptom on a scale from zero to three. The response scale presented to participants is of the following type:

0 I do not feel irritable
1 I feel irritable
2 I am irritable
3 I am so irritable I cannot stand it

BDI-II total scores are calculated by summing the ratings for the 21 items and range from zero to 63. Higher total scores reflect greater severity of symptoms. Participants are instructed to circle the number next to the statement that best describes the way they have been feeling over the past two-week period, including today. The duration of symptoms was extended in the BDI-II from one to two weeks consonant with DSM-IV diagnostic criteria and consequently appears to assess depression as a more persistent state or relatively stable syndrome.

Threshold Scores for Categories of Depression Severity

Beck et al. (1996) developed threshold score, or cut-scores, for categories of depression severity were using a sample of 127 outpatients from the University of Pennsylvania Medical School's Department of Psychiatry diagnosed according to the SCID-I (Spitzer, Williams, Gibbon & First, 1990). The description of the

clinical sample provides no information regarding demographic variables such as age, gender, race/ethnicity and education. Based on SCID-I results, patients were classified into four groups (non-depressed, mildly depressed, moderately depressed, and severely depressed). Of these groups, 44 patients were diagnosed as non-depressed based on the absence of any of the following: (1) Major Affective Disorder; (2) Depressive Disorder Not Otherwise Specified; (C) Dysthymic Disorder; And (D) Adjustment Disorder with Depressed Mood or Mixed Emotional Features. Fifty-seven patients were diagnosed with Major Depression-Single Episode or Major Depression-Recurrent with the DSM-III-R specifiers mild, moderate or severe. The diagnoses, if any, for the remaining 26 patients was not described in the BDI-II Manual (Beck et al., 1996). Optimal cut-scores were derived through examination of Receiver Operating Characteristic (ROC) curves and these curves, and cut-score guidelines are presented in the technical manual (Beck et al., 1996).

For the purpose of using the instrument as a screening instrument for Major Depression, the cut-score ranges place greater importance on the sensitivity (the probability of correctly classifying a respondent as depressed according to their BDI-II total score) of the test than the specificity (the probability of correctly classifying a respondent as not depressed) (Beck et al., 1996). These cut-scores, and within groups means and standard deviations are presented below in Table 7.

In research projects in which it is important to obtain as pure a group of individuals with depression as possible, Beck et al. (1996) suggest raising the cut-score to reduce the number of false positives and suggests that researchers use the ROC curves presented within the technical manual to establish the most appropriate cut-scores depending on the purpose of the assessment.

Table 7

BDI-II Threshold Scores for Categories of Depression Severity in Clinical Settings (N=101)

Total Score	Depression Severity	*M*	*SD*
0-13	No or minimal depression	7.65	5.9
14-19	Mild	19.14	5.7
20-28	Moderate	27.44	10.0
29-63	Severe	32.96	12.0

Reliability and Validity

Beck et al. (1996) used two samples to assess the psychometric properties of the BDI-II. One sample consisted of 500 psychiatric outpatients diagnosed with a variety of disorders recruited from four outpatient clinics. Of these 500 outpatients, 264 (53%) were diagnosed with mood disorders, 88 (18%) with anxiety disorders, 80 (16%) with adjustment disorders and 68 (14%) with a range of other disorders. The second sample consisted of 120 Canadian college students enrolled in an introductory psychology course. High internal consistency, with coefficient alphas of .92 and .93 was reported for the outpatient group and college student group respectively (Beck et al., 1996). While this may suggest some level of redundancy, when used to make clinical decisions about individuals, ideally, reliability coefficients are above .90 (Groth-Marnat, 2003).Test-retest reliability over a one week-period was .93 (Beck et al., 1996).

Numerous other research studies have examined the psychometric properties of the BDI-II using a variety of samples, translations, and administration formats, and when combined they provide strong support for its utility in disparate contexts. Among medical or psychiatric adolescent and adult samples, internal reliability coefficients for the BDI-II have been found to be greater than .88 (Buckley, Parker & Heggie, 2001; Coelho, Martins & Barros, 2002; Grothe et al., 2005; Kumar, Steer, Teitelman & Villacis, 2002; Osman, Kopper, Barrios, Gutierrez & Bagge, 2004; Sprinkle et al., 2002; Steer, Clark, Beck & Ranieri, 1999; Steer, Rismiller & Beck, 2000). Among college samples,

internal reliability coefficients have also been found to be high ranging from .84 to .93 (Al Musawi, 2001; Dozois, Dobson & Ahnberg, 1988; Schulenberg & Yutrzenka, 2001; Steer & Clark, 1997; Whisman, Perez & Ramel, 2000).

Evaluation of content, concurrent and discriminant validity has generally been favourable. The content of BDI-II items was derived from clinician consensus regarding symptoms of depressed patients combined with adherence to DSM-IV diagnostic criteria for depressive disorders. Concurrent validity is suggested by moderate to high correlations with clinical assessments of psychiatric patients (Beck et al., 1996). Additionally, moderate correlations have been found between the BDI-II and other indices of depression, including a correlation of .71 with the Hamilton Rating Scale for Depression (HRSD) (Hamilton, 1960) and .88 (Osman et al., 1997) with the DASS-D. Support for the BDI-II's ability to discriminate between anxiety and depressive disorders is supported in that BDI-II scores were more highly correlated with the depression dimension of other rating scales than the anxiety dimension. For example, Steer et al. (2000) found higher correlations between the BDI-II and the depression dimension of the Symptom Checklist-90-Revised (SCL-90-R) (.89) than with the anxiety dimension of the SCL-90-R (.71). Similarly, BDI-II scores were more highly correlated with the HRSD (.71) than with the Hamilton Rating Scale for Anxiety (HRSA) (.47). Importantly, the BDI-II has been able to discriminate psychiatric from non-psychiatric populations (Beck et al., 1996), and to discriminate the level of adjustment in psychiatric populations (Arnau, Meagher, Norris & Bramson, 2000; Beck et al., 1996).

In this thesis, perhaps the most salient investigations regarding the validity of the BDI-II relate to factor analytic studies. Factor analysis reveals the extent to which the inventory or questionnaire measures the factors that comprise the essence of the condition under investigation. In terms of depressive disorders, it is reasonable to suggest that factor analysis would reveal the cardinal signs, or key features of the disorder, these being depressed mood, and anhedonia (American Psychiatric Association, 1994). For the BDI-II, factor analyses conducted by Beck et al. (1996) indicate that the inventory comprises two factors. One is a Cognitive-Affective factor related to negative self-evaluative thoughts and depressive

feelings (e.g., feelings of worthlessness, suicidal thoughts and self-dislike) the other is a non-cognitive Somatic-Performance factor composed of the somatic symptoms and performance decrements of depression (e.g. crying, fatigue, and changes in appetite and sleep patterns).These two factors comprise the two subscales of the BDI-II and have been found consistently in differing samples including primary care medical patients (Arnau et al., 2001), clinically depressed outpatients (Steer et al., 1999), adolescents (Steer, Kumar, Ranieri & Beck, 1998), college students (Beck et al., 1996), geriatrics and geriatric inpatients (Steer, Rissmiller & Beck, 2000), However, Osman et al. (1997) found a slightly different three-factor structure (Negative Attitude, Performance Difficulty and Somatic Elements) using a sample of undergraduate college students.

Limitations of the BDI-II

Notwithstanding the generally favourable findings regarding the reliability and validity of the BDI-II, the validity of inventory as a measure of depression is contingent upon the validity of defining depression as a mental disorder comprising a variety of cognitive, affective and somatic symptoms. In this thesis, it is proposed that defining, and indeed measuring, depression in a manner consistent with the BDI-II is methodologically problematic.

Specifically, according to Beck and his colleagues (1996), the essential defining characteristic of depression is the Cognitive-Affective component and this is consistent with the cardinal signs of depression within the DSM nosology. However, in the BDI-II, items within the Cognitive-Affective subscale and Somatic-Performance subscale are equally weighted. That is, each item, or symptom, is rated by respondents in terms of intensity on a scale from zero to three, where zero reflects an absence of the symptom and three reflects a high level of the symptom. All items are then summed to form a depression scale total score and classified into categories of depression severity according to severity cut-scores. In this context, it may be possible for an individual with a physical illness to report high scores on items reflecting physiological symptoms while rating cognitive and affective items as zero. In this way, such an individual may score a depression scale total commensurate with a severity rating of moderate depression without endorsing any symptoms considered fundamental to any

definition of depression. Therefore, the inclusion of physiological symptoms has the potential to result in the artificial inflation of scores due to symptoms of illness, as opposed to depression.

Consistent with the above proposition, Ritterband and Spielberger (2001) found that the significantly higher BDI scores of cancer patients, relative to healthy control participants, were almost entirely due to elevated scores on the BDI Somatic-Performance subscale. Importantly, as there was no significant difference between the two groups in scores on the Cognitive-Affective subscale, the higher depression scale total scores of cancer patients were attributed to the physiological symptoms caused by their medical condition or treatment they were receiving and were not due to clinical depression. In line with Ritterband's findings, Callahan, Kaplan and Pincus (1991) found that the significantly higher depression scale total scores of patients with rheumatoid arthritis was primarily due to their elevated scores on the Somatic-Performance Subscale.

Taken together, the abovementioned findings have serious implications for researchers using the BDI-II to investigate depression in general population samples in which the physical health status of participant may not be assessed and therefore cannot be controlled. In such instances, the BDI-II depression scale total score may represent an inaccurate measure of depression severity.

Rationale for Scale Selection

The BDI-II was selected for use in the current research because it comprises the most widely used self-report measure of depression and has been found to be psychometrically sound. Further, as the aim of Study 2 is to investigate the comparative validity of self-report depression measures and SWB measures as novel measures of depression, the BDI-II offers an alternate conceptualisation of depression (psychobiological) to DASS-D and HADS-D (psychological) and therefore, may be directly compared with scales that operationalise to measure depression differently.

Depression, Anxiety and Stress Scales

The DASS (Lovibond & Lovibond, 1995a) are a set of three self-report scales designed to define and measure the core symptoms of depression, anxiety and stress. Importantly, the scales were constructed with the aim of developing maximum discrimination between the states of depression and anxiety. A commonly used measure, the DASS has received increasing attention in basic research (Brown, Chorpita & Barlow, 1998; Einstein, Lovibond & Gaston, 2000; Keogh & Cochrane, 2002) and in treatment outcomes research (Hooke & Page, 2002; Nieuwenhijusen, de Boer, Verbeek, Blonk & van Dijk, 2003; Norton & Hope, 2005; Norton, Hayes & Hope, 2004) due to its ease of administration. Through the development of the DASS, Lovibond and Lovibond aimed to further the process of defining, understanding and measuring the ubiquitous and clinically significant negative emotional states usually described as depression, anxiety and stress. In so doing, the authors of the scale aimed to satisfy the needs of both researchers and mental health professionals by providing them with a psychometrically sound measure that may be used reliably to isolate and identify specific elements of emotional disturbance (Lovibond & Lovibond, 1995b) within clinical and non-clinical samples.

Unlike other depression and anxiety scales, such as the BDI-II (Beck et al., 1996) and Beck Anxiety Inventory (BAI) (Beck and Steer, 1990), which have been found to contain overlapping symptoms anxiety and depression respectively (Lovibond & Lovibond, 1995b; Richter, Werner, Heerlein, Kraus & Sauer, 1998), the DASS was designed to assess the unique and unrelated aspects of depression and anxiety, as well as a stress factor. However, the development of the DASS was unconventional.

Scale Development

The DASS was constructed using an unusual research strategy. First, Lovibond and Lovibond collected data over the period 1979 to 1990, and excluding an initial item check with a patient sample, all major scale development was undertaken using over 30 normal non-clinical samples. Second, although clinical consensus was used to define factors initially, they were refined

empirically using a confirmatory factor analytic technique, multiple groups factor analysis (Harman, 1976), and a bootstrapping strategy, that the authors refer to as Simultaneous Multi-Scale Dimensioning (SMD). This strategy seeks to develop simultaneously scales designed to measure several empirically related but conceptually distinct dimensions, while at the same time achieving the maximum possible discrimination between the scales. In essence, the technique used by Lovibond and Lovibond begins with an *a priori* definition of the factors of interest, but permits bootstrapping to refine the initial factor definitions or to define new factors. Finally, in contrast to Beck and his colleagues (1996) who used previous versions of their depression inventory and diagnostic criteria from the DSM classificatory system to guide item selection, Lovibond and Lovibond, (1995a, 1995b), used no external criteria, arguing that existing depression and anxiety scales contained inconsistencies and overlapping symptoms that they were attempting to avoid.

DASS Constructs, Response Scale and Scoring

There are two versions of the DASS, the original 42-item self-report inventory (DASS42) and a half-length short-form version (DASS21) comprising a sub-set of items from the original 42-item DASS. Despite its shortened format, the DASS21 has been shown to be effective in discriminating between the three constructs of depression, anxiety and stress (Antony, Bieling, Cox, Enns & Swinson, 1998; Clara, Cox & Enns, 2001; Henry & Crawford, 2005; Lovibond & Lovibond 1995a; Norton, 2007). In this thesis, the depression scale of the DASS21 (DASS21-D) is used as a measure of self-reported depression in both the first and second study.

Like the DASS42, the DASS21 consists of three subscales (depression, anxiety and stress). Each of the subscales comprises seven items. The seven items of the DASS-D scale are psychological constructs and represent the cognitive and affective symptoms unique to depression. These items reflect the negative emotional states of anhedonia, inertia, hopelessness, dysphoria, lack of interest/involvement, self-deprecation, and devaluation of life. Item statements used to measure these states are presented in Table 8 below.

Table 8

DASS21-D Scale Items

Depression Symptom	Item
Anhedonia	I couldn't seem to experience any positive feeling at all
Inertia	I found it difficult to work up the initiative to do things
Hopelessness	I felt that I had nothing to look forward to
Dysphoria	I felt down-hearted and blue
Lack of interest/involvement	I was unable to become enthusiastic about anything
Self-deprecation	I felt I wasn't worth much as a person
Devaluation of life	I felt that life was meaningless

Each item is presented to respondents with the following response scale:

0 Did not apply to me at all

1 Applied to me to some degree, or some of the time

2 Applied to me very much, or a good part of the time

3 Applied to me very much, or most of the time

Participants are instructed to circle the number that best indicates how much each statement applied to them over the past week. The depression scale total score is calculated by summing the ratings for the seven items and multiplying this result by two. This calculation yields a minimum scale score of zero and a maximum scale score of 42 (Lovibond & Lovibond, 1995a). As the DASS21-D is a half-length version of the original DASS42 depression scale, doubling of the summed depression items allows for the direct comparison of scores to full-scale depression scores, and enables these to be interpreted by reference to the normative values for the full scale. Higher total scores reflect greater depression severity. This is because the DASS is based on a dimensional as opposed to categorical conceptualisation of psychological disorder. As such, Lovibond and Lovibond (1995a) assert that the DASS has no direct implications for the allocation of individuals to diagnostic categories. Nevertheless, in line with a dimensional approach, where the affective experience of normal individuals and

the clinically disturbed is essentially a difference of degrees, threshold scores for categories of depression severity are provided in the DASS Manual.

Threshold Scores for Categories of Depression Severity

Threshold scores, or cut-scores, for categories of depression severity were generated from the normative data set described in the DASS Manual (Lovibond & Lovibond, 1995a). This data set was based on six of the non-clinical samples used during scale development, comprising 1,044 males and 1,870 females (N=2,914) ranging in age from seventeen to 69 years. Participants were described as students from university classes in medicine, psychology and adult education, and employees from a railway workshop, a naval dockyard, an airline and a bank. No other demographic information was provided. Given that participants were either university students or employed adults from a narrow range of fields, it appears unlikely that the normative sample was demographically representative of the Australian general population. This potentially limits the external validity of the DASS normative data.

According to the normative data presented in the DASS Manual, the DASS-D mean score for women and men combined is 6.34 with a standard deviation of 6.97. Cut-scores, percentiles and corresponding *Z* scores for categories of depression severity are also outlined in the manual and are presented in Table 9.

Table 9

Categories of DASS Depression Severity

Depression Severity	Total Score	*Z* score	Percentile
Normal	0-9	<0.5	0-78
Mild	10-13	0.5-1.0	78-87
Moderate	14-20	1.0-2.0	97-95
Severe	21-27	2.0-3.0	95-98
Extremely Severe	28+	>3.0	98-100

Despite acknowledging that threshold scores were *based* on the normative data set, the precise procedure used for generating the threshold scores for the DASS-D and categories of depression severity in Table 9 are not provided in the DASS Manual. However, inspection of the distribution of scale scores, and corresponding percentiles, and *z* scores in Table 9 suggests that the cut-scores for each category of depression severity may have an empirical derivation. That is, scale scores for categories of depression severity appear to have been allocated on the basis of *z* scores and percentiles. Implementation of such an empirical strategy to generate threshold scores is in line with the empirically based research strategy employed by the authors of the scale to develop the DASS and conforms to the dimensional approach to classifying affective experience. Notwithstanding this, Lovibond and Lovibond (1995a) point out that DASS-D cut-scores and ratings of depression severity were later validated against number of clinical samples.

Reliability and Validity

Lovibond and Lovibond (1995a) used the six non-clinical samples described above to assess the psychometric properties of the DASS. Using these samples, the internal consistency of the DASS scales was found to be high with Cronbach's alpha coefficients of .91, .84, and .90 for the DASS respectively. The alpha values for the DASS-D scales were derived from one of the six non-clinical samples (N=717) and were found to be .81 for depression, .73 for anxiety, and .81 for stress, somewhat lower than the full DASS. However, a later study reported slightly improved psychometric properties for the DASS21 relative to the original DASS (Antony, Beiling, Cox, Enns & Swinson, 1998). Notwithstanding these differing findings, ideally, Cronbach's alpha coefficients for a scale are above .70 (DeVellis, 2003). In these terms, the DASS-D may be considered to have good internal reliability. Nevertheless, for clinical applications alpha coefficients of above .90 are preferable (Groth-Marnat, 2003), and in these instances, the full scales of the original DASS may have greater utility.

Numerous other research studies have shown that both the original DASS and DASS21 are reliable measures of depression, anxiety and stress in clinical and non-clinical adult samples (Antony et al., 1998; Brown, Chorpita, Korotisch & Barlow, 1997; Crawford & Henry, 2003; Clara, Cox & Enns, 2001; Duffy

Cunningham & Moore, 2005) and across administration formats (Shea, Tennant & Pallant, 2009) and in different cultural groups (Akin & Cetin, 2007; Daza, Novy, Stanley & Averill, 2002; Norton, 2007). Additionally, using principal components analysis and structural equation modelling, numerous studies have replicated the three scales and have accounted for up to 60% of the variance in a three-factor solution (Antony et al., 1998, Brown et al., 1997; Clara, Cox & Enns, 2001). In combination, extant data provide strong support for the reliability and validity of the DASS, and importantly for this study, the DASS21-D.

For example, among British general population adults, internal reliability coefficients for depression scale was .82 (Henry & Crawford 2005), and in both clinical and non-clinical adult samples have been found to be .94 (Antony et al., 1998). In depressed psychiatric samples reliability coefficients were also high (α = .92) and were high in geriatric primary care patients (α = .87) (Gloster et al., 2008), and in adolescent samples (α = .87) (Tully, Zajac & Venning, 2009). Additionally, the DASS21 has demonstrated a consistently stable and interpretable three-factor structure in many studies (Antony et al., 1998; Clara et al., 2001; Crawford & Henry, 2003). Notably, the depression scale has been found to represent anhedonic depression (Clara et al., 2001) and this is consistent with the cardinal signs of depression within both a categorical taxonomy and a dimensional understanding of affective experience. Finally, the DASS21 has demonstrated good convergent validity and discriminant validity. For example, the DASS21-D has been found to demonstrate high correlations with other indices of depression, for example r = .75 with BDI-II (Gloster et al., 2008) and r = .74 with BDI (Lovibond & Lovinbond, 1995b), and to be able to discriminate other negative emotional syndromes evidenced by lower correlations with anxiety and stress scales, for example r = .51 with BAI (Gloster et al., 2008), r = .54 with BAI, r = .54 DASS-A, and r = .56 with DASS-S (Lovibond & Lovibond, 1995b).

Most recently, the psychometric properties of the DASS21 were evaluated using Rasch analysis (Shea et al., 2009). In this study, the DASS21 was administered to 420 participants using two administration formats. Half of the sample completed a paper-and-pencil questionnaire, while the other half completed a web-based version of the scales. Shea and colleagues' results

supported the use of the 4-point response scale, the internal consistency reliability, and the unidimensionality of the three scales. Additionally, item functioning was found to be invariant to administration format, age, sex and education. However, in order to achieve satisfactory fit to the Rasch model, these researchers found that it was necessary to remove one item from each of the scales. For the depression scale, the mis-fitting item removed reflected the state of inertia ('I found it difficult to work up the initiative to do things'). Nevertheless, these researchers do not recommend the removal of DASS21 items in present research studies until further evidence accumulates to suggest the revision of scale content. In these terms, all DASS21-D scale items will be retained in the current research.

In summary, the DASS (Lovibond & Lovibond, 1995a) comprises three self-report scales designed to define and measure the core symptoms of depression, anxiety, and stress. The DASS21-D is a short form version of the full DASS-D, and conceptualises and measures depression as a psychological construct consistent with a dimensional understanding of affective experience. The DASS21-D is easily administered, comprising only seven items on 4-point response scale. Total scale scores may be referenced to the original DASS scale scores for classification into categories of depression severity ranging from normal to extremely severe. Additionally, the psychometric properties of the DASS21 have been extensively investigated using traditional methods, and most recently using Rasch Model analysis result from these studies provides strong support for the reliability and validity of the instrument, and therefore the DASS21-D may be viewed as psychometrically sound.

Rationale for Scale Selection

The DASS21-D was selected for use in Study 1 and Study 2 as it is predominantly a measure of the psychological symptoms of depression. As such, the DASS21-D may be more likely than other depression scales to measure the core construct of depression (i.e. a loss of positive affect). Additionally, as the DASS21-D measures depression essentially in terms of affective symptoms on a severity continuum, it is consistent with a dimensional conceptualisation of affective experience, and therefore in line with the framework for research in this thesis. Further, conceptualising depression in this manner allows the opportunity

to compare the DASS21-D directly with other depression scales that operationalise to measure depression differently (e.g., BDI-II or psychobiological symptoms). This is the aim of the Study 2 in the current research. This review will now turn to the examination of the final depression scale used in the current research, the HADS-D (Zigmond & Snaith, 1983).

Hospital Anxiety and Depression Scale

The HADS (Zigmond & Snaith, 1983) was developed more than 20 years ago as a brief self-report questionnaire to assist physicians and surgeons in the identification of depression and anxiety disorders among medical patients. Despite being developed for use in general hospital and medical outpatient clinics, the HADS is also widely used as a screening tool for depression and anxiety in general population samples (Crawford, Henry, Crombie & Taylor, 2001).

The HADS comprises two seven-item subscales, representing depression and anxiety. Scale items were selected to reflect the psychological symptoms of the respective constructs. Physiological symptoms, such as dizziness and headaches, which could be related to the physical wellbeing of the respondent, were removed from the scales by Zigmond and Snaith (1983) in an attempt to differentiate the effects of physical illness from mood disorders. Additionally, to ensure that items discriminated between depression and anxiety, the authors of the scale removed any items that may be related to both constructs (e.g., fatigue). In this context, depression items were selected to reflect only the psychic symptoms of depression, specifically, the anhedonic state as this is likely the core psychopathological feature of depression. As such, the depression subscale of the HADS (HADS-D) and DASS-D scales appear to operationalise depression in a similar manner, and differently to the BDI-II.

In the current research, only the HADS-D is used. Therefore, presentation of the scale content, methods for calculating scale total scores and review of empirical findings related to the reliability and validity of the HADS will be limited to the HADS-D scale.

Depression Scale Items, Response Scale and Scoring

The HADS-D scale items are each presented to respondents with a 4-point response scale (zero to three). Respondents are asked to select the statement that comes closest to the way they have been feeling over the past week. Depression items and their corresponding response choices are presented in Table 10.

Table 10

HADS Depression Items and Response Choices

Item		Response Choices			
1	I still enjoy the things I used to enjoy	Definitely as much	Not quite so much	Only a little	Hardly at all
2	I can laugh and see the funny side of things	As much as I always could	Not quite so much now	Definitely not so much now	Not at all
3	I feel cheerful	Not at all	Not often	Sometimes	Most of the time
4	I feel as if I am slowed down	Nearly all the time	Very often	Sometimes	Not at all
5	I have lost interest in my appearance	Definitely	I don't take so much care as I should	I may not take quite as much care	I take just as much care as ever
6	I look forward with enjoyment to things	As much as I ever did	Rather less than I used to	Definitely less than I used to	Hardly at all
7	I can enjoy a good book or radio or TV programme	Often	Sometimes	Not often	Very seldom

The response choices in Table 10 differ for each of the seven items although all items are scored from zero to three, where zero reflects an absence or low incidence of the depressive state being measured and three reflects a higher incidence of that state, or absence of positive emotional experiences over the past week. Depression scale totals are calculated by summing the seven item scores; minimum score = 0, maximum score = 21. Higher scores reflect greater depression severity. Notably, the five of the seven HADS-D items are framed in the positive, whereas all DASS21-D items measure the psychological symptoms

of depression through negative constructs. For example, the HADS-D operationalises to measure anhedonia, through positively worded statements, ‘I feel cheerful’, and ‘I can laugh and see the funny side of things’. The response choice ‘Not at all’ corresponds to an item score of three reflecting greater symptom severity for these items. Contrastingly, the DASS21-D measures anhedonia through a negatively worded item ‘I couldn’t seem to experience any positive feeling at all’. In these terms, it is interesting to speculate whether the repetition of negative items may impact, or have a deleterious effect on the affective tone of the respondent, and if so, whether this may artificially inflate DASS21-D total scores relative to HADS-D total scores.

Threshold or Cut-scores for Cases of Depression

Zigmond and Snaith (1983) developed cut-scores for cases of depression using a sample of 98 patients from general medical outpatient clinics. The sample is described as comprising both men and women aged from sixteen to 65 years who suffered from a wide variety of illnesses and complaints. No other information is provided.

Patients completed the depression and anxiety scales and were interviewed by the authors of the scale who rated the overall severity of depression and anxiety of each patient using a set of standard questions. Depression was assessed through the questions: ‘do you take as much interest in things as you used to?’; ‘do you feel cheerful?’; ‘do you laugh as readily?’; and ‘do you feel optimistic about the future?’ Psychiatric interviews lasted 20 minutes for each patient. Interviewers were blind to the outcome of the scores on the self-assessment questionnaire. Results from psychiatric assessment and scores on the self-report scales were then compared. Zigmond and Snaith found that the same scale score ranges produced the best fit with severity ratings awarded through psychiatric assessment for cases of depression and anxiety.

For the depression subscale, scores of seven or less identified non-cases, eight to ten doubtful cases, or just suggestive of the presence of the depressive state, and 11 or greater, definite cases. These cut-scores for depression produced only one false positive, no false negatives, five borderline scores among the non-

cases and one borderline score among the definite cases. Additionally, examination of Spearman correlations of the HADS-D scores and psychiatric ratings of severity were $r = .70$. These results suggested that scale scores could also be used as indicators of depression severity. Despite these findings, a review of more than 200 studies that have investigated the psychometric properties of the HADS suggests that there is no single, generally accepted cut-score used in interpreting the HADS scales (Hermann, 1997). However, an updated review of 747 studies that have examined the validity of the HADS indicates that in most studies, the optimal balance between sensitivity and specificity was achieved when caseness was defined by a score of eight or greater (Bjelland, Dahl, Haug & Neckelmann, 2002). Based on this latter finding, HADS-D scale total scores of zero to seven will reflect non-cases and scores of eight to 21 will reflect cases of depression in the current research.

Reliability and Validity

Zigmond and Snaith (1983) investigated the internal consistency, convergent and discriminant validity of HADS using a sub-set (N=50) of the total patient sample described previously. Item-total correlations were all significant ($p < 0.02$) ranging from .6 to .3. Additionally, concurrent and discriminant validity was suggested by higher correlations between the HADS-D and clinician ratings of depression ($r = .79$) and lower correlations with clinician ratings of anxiety ($r = .08$) respectively.

Numerous other research studies have examined the psychometric properties of the HADS in various contexts, using a number of different translations and administration formats, and together provide support for its utility in a variety of settings. The data on the reliability and validity of the HADS from most of these studies has been summarised in two extensive reviews 200 research papers (Hermann, 1997) and from 71 studies (Bjelland et al., 2002). The collated findings within these reviews provide support for the internal consistency, concurrent, and discriminant validity of the HADS in medical, psychiatric and general population samples. Specifically, internal consistency (Cronbach's alpha) coefficients for various versions of the HADS (Arabic, Chinese, Dutch, English, French, German, Portuguese, Spanish and Swedish) range from .67 to .90 (mean =

.82) for the depression scale (Bjelland et al., 2002; Hermann, 1997). Additionally, using the German version of the HADS (N=6,200), test-retest reliability for the HADS-D over a two-week period was found to be high; r = .85. In these terms, the HADS may be viewed as reliable using numerous translations of the instrument.

The concurrent validity of the HADS-D is suggested by moderate to high correlations with other indices of depression, including a correlation of .71 with the BDI (Lisspers, Nygren & Soderman, 1997), .80 with the Montgomery-Asperg Depression Rating Scale (Uphadyaya & Stanley, 1993), and .69 with the Symptom Checklist-90 scale (SCL-90) (Depression) (Spinhoven and van der Does, 1997). Additionally, using cut-score of 8+, the HADS-D has demonstrated good to excellent sensitivity and specificity (0.70 to 0.90) (Bjelland et al., 2002) in detecting cases of depression diagnosed according to a variety of 'gold standard' diagnostic interviews.

For factorial validity, the majority of factor analytic studies reviewed in Bjelland et al.'s (2002) paper used principal components analyses and demonstrated a two-factor solution in accordance with the depression and anxiety scales of the HADS respectively. Further, a two-factor solution was found to explain more than 50% of the variance in a large sample of patients with cancer (Moorey et al., 1991). Importantly, in two general population studies, a two-factor solution was found to be stable across age and gender (Spinhoven et al., 1997; Lisspers, Nygren & Soderman, 1997).

Notwithstanding the above findings, some researchers have questioned whether the HADS scales measure two separate aspects of depression and anxiety, comprise scales that contain significant overlap (Chaturvedi, 1991), or represent three factors, as found previously (Dunbar, Ford, Hunt & Der, 2000; Martin & Newell, 2004; Martin, 2005) Certainly, correlations between the depression and anxiety scales have been found to be high; ranging from a low of .49 to a high of .74 (Hermann, 1997; Bjelland et al., 2002). In these terms, some researchers have argued for the use of the HADS total score as a single measure of general distress (Martin, Tweed & Metcalfe, 2004). This proposition has been supported in a recent Rasch analysis that assessed the appropriateness of using a HADS total

score as a measure of general distress in an outpatient sample (*N*= 296) (Pallant & Tennant, 2007). The results of the Rasch analysis, and those of other studies that have demonstrated a three-factor solution for the HADS (Dunbar et al., 2000, Martin & Newell, 2004; Martin, 2005), raise serious concerns over the validity of the HADS anxiety scale. Specifically, these researchers have found that it is the anxiety subscale that splits into two creating a three-factor solution, and the best fit to the Rasch model, whereas the depression scale remains intact. While these results have very important implications for the HADS as a measure of both anxiety and depression, they do not impact the validity or integrity of the HADS-D scale. Thus, these results, and those described throughout this section, provide strong support for the reliability and validity of the HADS-D and consequently, support its use as valid measure of depression in the current research.

Rationale for Scale Selection

The HADS-D was selected for use as a measure of depression in Study 2 as, like the DASS21-D, it is a measure of the core symptoms of depression, and these are psychological in nature. However, unlike the DASS21-D, the HADS-D measures these symptoms, predominantly the anhedonic state, through positive item statements. Here, instead of casting items in the negative, such as 'I do not feel cheerful', five of the seven HADS-D items are phrased positively, for example, 'I feel cheerful'. In these terms, although the DASS21-D and HADS-D conceptualise and measure depression as a psychological construct, they operationalise to measure the psychological symptoms of depression in a contrasting manner. Thus, the inclusion of the HADS-D in Study 2, affords an opportunity to compare HADS-D depression total scores with those of the DASS21-D and BDI-II using the same sample.

CHAPTER 3: SUBJECTIVE WELLBEING AND DEPRESSION

This chapter presents and reviews findings from research studies that have investigated the relationship between LS, SWB and depression as evidence that supports the proposition that depression may be conceptualised as suppressed HPMood, subsequent to the failure of homeostasis. Following this, the precise predictions and diagnostic approximation made within SWB homeostasis theory (Cummins et al., 2009, in press) regarding the relationship between SWB that allow for the testing of homeostasis as the proposed mechanism of SWB, and the examination of depression in terms of the loss of the normal levels of SWB are described. Finally, this chapter will conclude with the aims of Study 1 and the hypotheses regarding the relationship between SWB and depression to be tested therein.

Evidence for the Relationship between Subjective Wellbeing and Depression

It has been argued that depression may be conceptualised as a loss of normal levels of SWB, or more precisely the suppression of HPMood, subsequent to the failure of homeostasis (Cummins & Lau, 2003; Cummins et al., 2002). This argument appears meritorious, given that a lack of LS, the fundamental construct measured in SWB scales, is a common element of depression (Lewisohn, Redner & Seeley, 1991). Further, SWB is essentially a measure of positive affectivity, and depression is differentiated from other disorders associated with high levels of negative affect by a relative absence of positive affect (Kendall & Watson, 1989; Watson & Clark, 1991; Watson & Tellegen, 1985). In these terms, it is not surprising that empirical studies have consistently demonstrated that SWB and depression are related.

The relationship between SWB and depression has been investigated in a number of ways that, in combination, provide general support for the proposition that SWB and depression may represent the same underlying construct. Specifically, studies have consistently found that there is a moderate to strong inverse relationship between depression and LS (Cheung & Bagley, 1998; Heady, Kelly & Wearing, 1993; Hong & Giannakopoulos, 1994; Lewis, Dorahy & Schumaker, 1999; Simpson, Schumaker, Dorahy & Sarvagya, 1996), that levels of

SWB are reduced for those with high levels of depressive symptomatology (Bonciatto, Dew, Zaratiegui, Lorenzo & Pecina, 2001, Broe et al., 1998; Heady et al., 1993; Lewinsohn, Hoberman & Rosenbaum, 1988; Jho, 2001; De Leval, 1999; Lewinsohn, Redner & Seely, 1991; Ritzner, Kurs, Kostizky, Ponizovksky, Modai, 2002), that SWB increases as depressive symptoms decrease (Hansson, 2002; Lewinsohn et al., 1991), and that SWB is a significant predictor of depression (Abbey & Andrews, 1984; Hong & Giannokopoulos, 1994; Lewinsohn et al., 1991). Additionally, several studies have demonstrated SWB levels are lowest for people with depressive disorders relative to people with other psychiatric disorders (Hansson, 2002; Heady et al., 1993; Russo et al., 1997), and relative to people with chronic medical illnesses (Bonciatto et al., 2001; Broe et al., 1998).l

In addition to the above findings, other studies have suggested that measures of SWB may be used to identify individuals at risk of developing depression (Koivuuma-Honkanen, Kaprio, Honkanen, Viinamäki & Koskenvuo, 2004; Koivuuma-Honkannen et al., 2001; Lewinsohn et al., 1991). For example, Koivumaa-Honkanen et al. (2004) investigated the cross-sectional and longitudinal relationship between LS and depressive symptoms. In this study, over 9,676 healthy Finnish adults who had responded to postal questionnaires in 1975, 1981 and 1990 were surveyed. LS was measured using a four-item scale that had been adapted from a QoL questionnaire designed for use in Nordic countries. This scale was included at all three time points, and in 1990, the BDI was included in the questionnaire. These researchers found a strong linear association between concurrent LS and depression measures (r = 0.6). Importantly, using ROC techniques the accuracy of the LS scale to detect cases of moderate-severe BDI depression was found to be 94% overall, with a sensitivity of 87% and specificity of 88% respectively using a LS cut-score of 11/12. Together, these findings suggest a strong association between LS and depression, and suggest that measures of LS may have utility as indicators of depression.

From a longitudinal perspective, Koivumma-Honkanen and colleagues (2004) found that people with lower levels of LS in 1990 relative to their satisfaction with life in 1980, and 1975, demonstrated an increased risk of

moderate-severe depression (OR = 10.4, 95% CI = 6.1-17.6; OR = 6.7, 95% CI = 4.2-10.9 respectively). Importantly, these results corroborate the findings from an earlier study conducted by Lewinsohn et al. (1991) in which low levels of LS were also found to be a significant risk factor for future depression. In this context, these researchers concluded that LS scales could identify a group of individuals from the general population at high risk of developing depression. Further, that as low levels of LS had been found to predict other adverse health outcomes such as suicide (Koivumaa-Honkanen et al., 2001), these researchers recommended that SWB measures should be routinely included in research questionnaires and clinical practice. In combination, these results appear to provide preliminary support for the utility of SWB measures as screening instruments for depression risk in general population samples. However, this extensive study is not without limitations.

Inspection of the four items comprising the LS scale used in Koivumma-Honknanen and colleagues research reveals that it is comprised of a group of disparate items that in combination yield a scale with limited interpretability. Specifically, participants are asked 'do you feel that your life at present is …' (1) interesting, fairly interesting, fairly boring, or very boring; (2) very happy, fairly happy, fairly unhappy, or very unhappy; (3) very easy, fairly easy, fairly hard, or very hard'. The final item asks respondents whether, at present, they are (4) very lonely, fairly lonely, or not at all lonely. Measuring LS using these items is problematic for several reasons.

First, items are cast in the form of 'the present'. This has the potential to elicit responses based on momentary emotions as opposed to stable enduring evaluations of LS based on HPMood. As such, responses may be influenced by transient affective states and therefore may not represent true measures of satisfaction with life as a whole. Second, the items appear confounded. For example, responses that life is 'interesting' (item 1) and 'very easy' (item 3) are scored as reflecting the highest rating of LS. However, it is possible that some individuals with high levels of LS may equate a 'very easy' life (item 3) with a 'very boring' life (item 1), yet 'very boring' is scored as reflecting very low levels of LS. Finally, there are no psychometric data reported for the modified scale used

to measure LS. In combination, these factors raise serious questions of measurement reliability and limit the interpretability of Koivumma-Honkanen's findings. Notwithstanding these methodological difficulties, the growing body of research that has investigated the relationship between SWB, LS and depression are undoubtedly related.

Subjective Wellbeing Homeostasis and Depression

According to theory (Cummins & Nistico; Cummins et al., 2009, in press; Cummins, 2009, in press), SWB is actively managed by a system of genetically pre-wired, neurological systems that have evolved for this purpose. It is proposed that this system is consistent with the concept of homeostasis. Consequently, describing SWB management in terms of homeostasis gave rise to clear theoretical predictions regarding the relationship between SWB and other variables if the system is to be considered homeostatic. These were presented in Chapter 1. In line with these predictions, Cummins et al. (2009, in press) has proposed that SWB will demonstrate an inverse curvilinear relationship with increasing levels of negative challenge. These researchers cite evidence for this proposal from an earlier investigation into SWB and depression conducted by Davern (2003, unpublished). These cited results show that SWB demonstrated a curvilinear relationship with depression across increasing depression scores, and homeostatic failure was found to correspond to a depression rating of moderate using the DASS. In this context, depression was seen to represent the challenging agent, with low depression scores reflecting a low level of homeostatic challenge and high depression scores reflecting a very strong challenge. Thus, Cummins and colleagues interpreted Davern's results as consistent homeostasis theory and consistent with Figure 3 presented below.

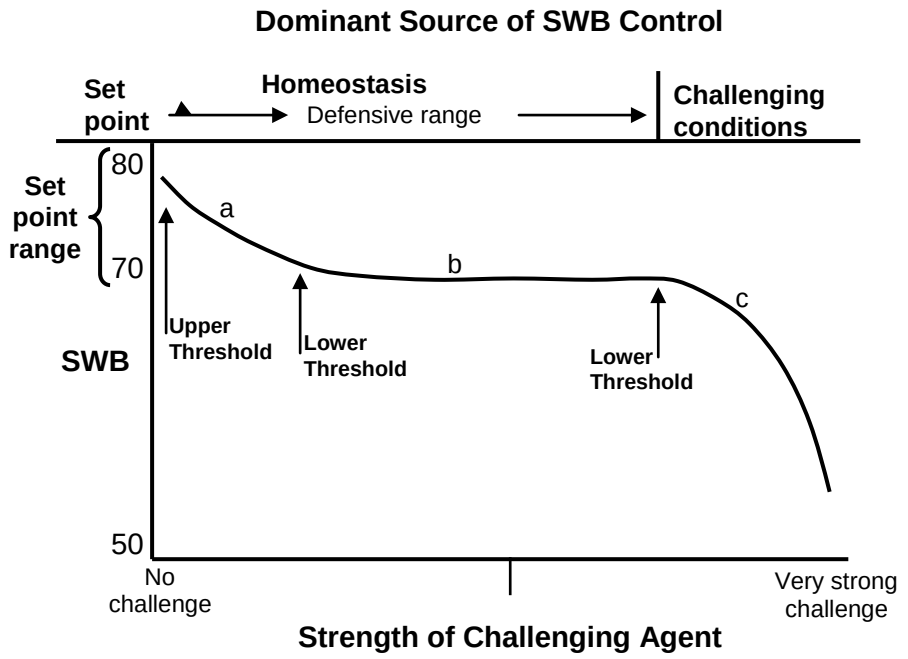


Figure 3. Changing levels of SWB as homeostasis is challenged (Cummins, Lau & Davern, 2009, in press).

Figure 3 depicts the changing levels of SWB across increasing strength of a challenging agent. In this instance, increasing negative challenge may be taken to reflect increasing depression severity. According to Cummins and colleagues (2009, in press) under conditions of no challenge, or threat, SWB will exhibit its set-point of 75 points on average (Cummins, 2003). As sources of challenge and support are experienced, it is proposed that the level of SWB will vary within its set-point range. This is phase 'a' in Figure 3. The explanation Cummins (2009, in press) offers for this is because, at any moment, where the actual SWB is located within the set-point range is a probability statement determined by the balance of challenge and support. As challenge is applied to the homeostatic system, the probability increases that SWB will inhabit the lower reaches of the normative range. At this point, SWB increasingly becomes bonded to the lower threshold where it appears to hold steady demonstrating a plateau effect.

Then, as the strength of the challenging agent continues to intensify, the homeostatic system responds by increasing the strength of defence. Here (phase 'b' in Figure 3), homeostasis functions to buoy SWB levels despite increasing

strength of challenge. Hence, SWB will appear to 'hold-the-line' across the lower threshold of the set-point range (70 points) (Cummins, 2003) demonstrating a 'plateau effect'. However, at some intensity of challenge, homeostasis will be overwhelmed, and will fail. This corresponds to phase 'c' in Figure 3. Once homeostasis fails, SWB levels will fall sharply indicating that homeostasis is no longer the dominant source of SWB control, rather, SWB control has passed to the challenging agent. In Davern's (2003, unpublished) investigations, homeostatic failure (SWB < 70 points) was found to correspond to a DASS depression rating of moderate.

While it is encouraging that Davern's findings conform to Figure 3, it is important to note that the number of cases per group was small (N<20), raising questions of reliability. In these terms, whether SWB and depression evidence a curvilinear relationship consistent with the operation of homeostatic processes requires verification. This investigation will be undertaken in Study 1.

In addition to the demonstration of a curvilinear relationship between SWB and depression, Cummins has proposed a number of diagnostic approximations for SWB scores. This guide for interpreting SWB scores is based on the assumption that each individual has a genetically determined set-point range that lies somewhere between 55 and 95 points, or 75 points on average. These values are derived from the extensive studies that established the normative ranges for SWB reviewed in Chapter 1. The guidelines for interpreting SWB scores based on group mean score proposed by Cummins, Lau & Davern (2008, unpublished) are as follows: 70+ points = normal; 51-69 points = either a low set-point or strong homeostatic challenge, perhaps even defeat for those individuals with high set-points; less than 50 points = homeostatic defeat and depression. These diagnostic approximations will be investigated in Study 1.

As the global incidence of depressive disorders is predicted to escalate rapidly over the coming decade, (WHO, 2007), and if depression represents the suppression of HPMood following homeostatic failure, then understanding the mechanisms through which SWB may be controlled and maintained over time would appear critical. It is argued that such understanding may pave the way for a more parsimonious conceptualisation of depression, and allow for the

development of interventions aimed at enhancing mental health and preventing mental ill health, rather than ameliorating the symptoms of depression only after they have emerged. A fundamental step towards elucidating the mechanisms that underpin the maintenance of SWB across the lifespan is the systematic testing of the predictions derived from Cummins' theory of SWB homeostasis. In this way, the validity of a homeostatic operating system as the proposed controller of SWB may be established.

However it must be noted that while it is appropriate to focus only on Cummins' model for the purpose of achieving the aims of this project, there are other models of affect, such as Carver and Schier's (1982) control theory of affect, and theories of mood management, for example Larsen's (2000) cybernetic control theory of mood management, that may have relevance in the context of SWB and Depression and could be examined in future research in this area.

Summary and Study 1 Aims

Extant empirical evidence supports the notion that depression and SWB are inversely related. This is in accordance with the theoretical proposition that depression may be conceptualised as a loss of normal levels of SWB subsequent to the failure of homeostasis (Cummins, 2009, in press; Cummins et al., 2009, in press). Additionally, data linking low levels of LS with an elevated long-term risk for the development of depression and suicide may provide reserved support for the utility of SWB measures as screening instruments for depression risk in general population samples. Importantly, these findings are consistent with a dimensional understanding of affective experience. In this context, it is proposed that SWB as a measure of affective normality, and depression, as the loss of the normal positive sense of SWB, or more precisely HPMood, may comprise the end-points on a continuum of positive affect. This proposition is in line with the predictions made within SWB homeostasis theory regarding the relationship between SWB and depression.

Cummins and Nistico (2002) have proposed that SWB is actively managed by a system of dispositional, genetically pre-wired, neurological systems that have evolved for this purpose .The operation of this management system is proposed to

be consistent with the concept of homeostasis and forms the basis of SWB homeostasis theory. Consequently, describing SWB management in these terms gave rise to very clear theoretical predictions regarding the relationship between SWB and other variables and affords an opportunity to test homeostasis as the proposed mechanism of SWB control in this thesis. Additionally, these researchers have proposed a guide to interpreting SWB scores under increasing conditions of challenge. Specifically, they have made predictions regarding the likelihood of depression as a function of these scores. These theoretical predictions and diagnostic approximations germane to investigations undertaken in this study are summarised below.

1. There must be a threshold value for SWB that is being defended by homeostatic processes.
2. There must be evidence that as the threshold value is approached the homeostatic system works harder than normal to retain control. In this context, SWB levels remain relatively stable and demonstrate a 'plateau' effect despite conditions of increasing challenge.
3. Once the threshold value has been exceeded, there must be evidence that homeostasis has failed and is no longer controlling SWB.
4. If the individual SWB score lies above 70 points the homeostatic system is likely to be functioning normally and the person is not depressed.
5. If the individual score lies below 50 points, the person is highly likely to be depressed.

The aim of this study is to investigate the above propositions made within the context of SWB homeostasis theory (Cummins, 2009, in press). In this study, SWB is measured by the Personal Wellbeing Index-Adult Version (PWI-A) (International Wellbeing Group, 2006) and depression is measured by the DASS21-D (Lovibond & Lovibond, 1995).

Study 1 Hypotheses

To address these research aims, a number of hypotheses are proposed regarding the relationship between SWB and depression. These hypotheses are that:

1. Consistent with previous empirical research and SWB homeostasis theory, depression will demonstrate moderate negative correlations with SWB.
2. If the individual SWB score lies above 70 points using the PWI-A, the person will not demonstrate a rating on the DASS consistent with depression.
3. All individuals with a SWB score of less than 50 points will demonstrate a rating on the DASS consistent with depression.
4. In confirmation of the operation of homeostatic processes to maintain normal levels of SWB, SWB scores will demonstrate a curvilinear relationship with increasing depression scores.

The first three hypotheses will be tested in the following manner. First, to establish the validity of the DASS21-D, the psychometric properties of the DASS21-D and the comparability of data drawn from surveys using two versions of the DASS21-D (4-point response scale *versus* end-defined 11-point response scale) are examined. These investigations will include an examination of the means and standard deviations for the DASS21-D and PWI-A, and assessment of the correlations between the DASS21-D and the PWI-A in eight separate longitudinal cohorts. Following this, the diagnostic approximations made within SWB homeostasis theory regarding the occurrence of depression in the context of various levels of SWB will be tested. These tests will be performed through an examination of the distribution of DASS21-D scores within specific PWI-A score categories. These latter tests will also afford the opportunity to evaluate the differential manner in which cases are classified according to two versions of the DASS21-D (4-point and 11-point).

The second set of analyses will present results from a fine grain analysis of the relationship between SWB and depression and relates to testing the fourth

hypothesis. Namely, principal components analysis of the DASS21-D will initially be performed to confirm the structure of the depression scale. Following this, standard multiple regression will be used to examine the contribution of DASS21-D items to the prediction of SWB as measured by the PWI-A. Then, means and standard deviations for the PWI-A, DASS21-D total score, and DASS21-D items, will be evaluated overall, and within ratings of DASS21-D depression severity. SWB means will then be mapped across small equivalent increments of scores to assess whether SWB demonstrates an inverse curvilinear relationship with depression. Finally, to investigate the presence of a significant plateau in SWB levels across the lower homeostatic threshold of 70 points, SWB means will be sequentially tested using analysis of variance techniques.

CHAPTER 4: METHOD

Sample

All data for analyses in this thesis are drawn from the Australian Unity Wellbeing Index (AUWI). This is an ongoing research project arising from a partnership between Australian Unity and the Australian Centre on Quality of Life at Deakin University. Through quarterly cross-sectional telephone surveys and subsequent longitudinal surveys completed in written form, the AUWI comprises a subjective index of Australians' satisfaction with their lives and their life in Australia. The project commenced in April 2001 and to date, sixteen cross-sectional and eighteen longitudinal surveys have been conducted. Each cross-sectional survey comprises a new demographically representative sample of approximately 2,000 Australians. It involves a telephone interview containing a standard set of demographic questions, questions that measure SWB (PWI-A) and life within Australia (National Wellbeing Index), and other survey specific items. At the conclusion of these interviews, participants are asked if they would like to be involved in future surveys through written questionnaires. Those who indicate that they would like to remain involved provide contact details for mailing purposes. Approximately three months later, participants are mailed a written questionnaire and are followed up with additional questionnaires annually thereafter. Written questionnaires contain approximately 100 items that measure personal, neighbourhood and national wellbeing, the occurrence and impact of

recent life events, and contain validated test instruments that measure constructs such as depression, anxiety, stress, self-esteem and perceived control.

Longitudinal Participants

The longitudinal sample participants were drawn from cross-sectional Surveys 2-14 conducted between September 2001 and October 2005. Participants had been followed-up for various periods as detailed below.

Of the 25,095 participants in these cross-sectional surveys, 71% indicated that they would like to participate in future surveys and were sent a follow-up written questionnaire. A total of 5,482 people (response rate = 30.8%) returned the written questionnaire. These participants comprise the second wave of the project and were identified as the longitudinal cohort, comprising 37.2% males and 62.3% females with a mean age of 52.3 years (*SD* = 15.8). The people who returned their survey were sent further written questionnaires at intervals of about 12 months and their returned questionnaires constitute the third (*N*=2521), fourth (*N*=1355), and fifth (*N*=548) waves of the project.

Sample for Initial Analyses

The original 4-point response scale of the DASS was used to measure depression in three longitudinal surveys. These surveys comprised participants completing their first and second longitudinal survey.

In another twelve longitudinal surveys, DASS21 items were presented to participants using the 11-point response scale. However, in five of these surveys, there were systematically missing data for the depression scale due to an administrative error. This error resulted in the omission of three of the seven depression scale items from questionnaires. These surveys were subsequently excluded from analyses. In an additional survey, a problem with participants' unique identification codes resulted in exclusion of all data from further analyses. Hence, the total number of surveys was reduced from twelve to six and comprised participants' responses from time points second, third and fourth waves of the project.

As the samples comprised repeated measures data, only primary cases (data from an individual's first longitudinal survey) were retained in the data set. This resulted in a further reduction in sample size. Hence, the sample using the 4-point response choice comprises 1,294 participants and the sample using the 11-point end-defined response scale comprises 2,371 participants.

Measures

Global Life Satisfaction Item

GLS represents the most general, abstract, overall rating of LS. It is most commonly measured through the question, 'how satisfied are you with your life as a whole?' In the AUWI surveys, the response format is an 11-point end-defined scale anchored by zero to ten with a mid-point of five, where 0 = very dissatisfied, 5 = neutral, and 10 = very satisfied. Research has demonstrated that this measure is reliable (Larsen, Diener & Emmons, 1985), exhibits moderate stability and appropriate sensitivity to changing life circumstances (Eid & Diener, 2004) and has been found to be particularly consistent in Western countries (Cummins, 1995, 1998, 2003). While GLS is an excellent measure of SWB, it is not as reliable as multi-item measures (Cummins, 2003).

The Personal Wellbeing Index

The PWI-A (International Wellbeing Group, 2006) represents a first-level deconstruction of satisfaction with life as a whole. It adopts a domain-level approach in which each item measures satisfaction with a different life area, and in which each domain explains unique variance in GLS.

The PWI-A has evolved from the ComQoL (Cummins, McCabe, Romeo & Gullone, 1994), and has been refined to its current form by the International Wellbeing Group (2006). To date the index is used by more than 100 researchers in 50 countries and provinces, with an objective of making the PWI-A a valid cross-cultural instrument (International Wellbeing Group, 2006).

The PWI-A comprises eight items of satisfaction that correspond to different QoL domains. These domains are: standard of living; health; achieving

in life; relationships; safety; community-connectedness; future security; and religiosity/spirituality. The domain of spirituality/religion, added to the PWI-A in November 2006, is not included in PWI-A scale used in the first study in this thesis.

Satisfaction with each of the life domains is measured individually. Respondents are asked to rate their satisfaction on an end-defined 11-point scale anchored by zero (very dissatisfied) and ten (very satisfied) with a neutral mid-point of five. Scores for the seven domains are averaged to yield the PWI-A score, which is then converted to a percentage point scale from zero to 100.

The PWI-A domains were validated by regressing them against satisfaction with life as a whole. In Australian samples, the largest amount of unique variance is typically contributed by the domain of standard of living, followed by personal relationships and achievements in life respectively. The domain of safety does not make a unique contribution in the Australian samples. However, the domain of safety has been retained in the PWI-A as data from other countries indicate that does make a unique contribution in other cultures (International Wellbeing Group, 2006).

The PWI-A has sound psychometric properties. A maximum variation of 3.2 percentage points in SWB has been found between the sixteen surveys of the Australian population to date (Cummins, 2006). Cronbach's alpha for Australian samples lies between .71 and .85, and test-retest reliability over a one to two-week period has demonstrated an intra-class correlation coefficient of .84 (Cummins & Lau, 2005).

Depression, Anxiety and Stress Scales

As the DASS was reviewed in Chapter 2, only a brief overview will be provided here. Lovibond and Lovibond (1995a) designed the DASS as a self-report measure of depression, anxiety and stress with the aim of allowing maximum discrimination between the three constructs, most notably, between the constructs of depression and anxiety.

Depression was measured using the DASS21 in this study. The DASS21 is a half-length version of the original 42-item DASS, and despite its shortened format, has been shown to be effective in discriminating between the three constructs (Antony, Bieling, Cox, Enns & Swinson, 1998; Lovibond & Lovibond, 1995). The DASS has sound reliability. Alpha values for the three DASS subscales are .91, .84, and .90 for depression, anxiety and stress respectively. For the DASS21 7-item scales, alpha values are: depression 0.81; anxiety 0.73; and stress 0.81. (For depression scale content, 4-point response choices, and scoring see Chapter 2.)

As two versions of the DASS21-D, one using the original 4-point response format and the other an 11-point end-defined response scale anchored by zero and ten (Jones & Thurstone, 1955) are used in initial analyses, the rationale for using an 11-point scale is outlined below.

Rationale for using 11-point end-defined response scales.

The 11-point end-defined response scale was presented to participants based on previous research conducted by Cummins and Gullone (2000). These researchers argued that Likert-type scales commonly employed to measure psychological variables, such as SWB, are not sufficiently sensitive for using these variables as measures of outcome. On reviewing the literature, Cummins and Gullone found that increasing the number of choice-points beyond five or seven did not systematically compromise scale reliability, although such expansion did enhance scale sensitivity. Further, these researchers suggested that naming all response choice-points detracts from the interval nature of the data being measured. As such, a solution to these issues was proposed through adoption of an 11-point end-defined scale (Jones & Thurstone, 1955) that offered a rating from zero to ten. This range was suggested to lie within the range of common experience and, therefore, could be easily utilised as a rating system (Cummins & Gullone, 2000).

Lovibond and Lovibond (1995) found that depression and the states measured by DASS depression items using the DASS demonstrated a smooth progression across ratings of depression severity. Consequently, these researchers

concluded that depression and the states of depression appear to exist on a continuum, and as such, support a dimensional as opposed to categorical view of depression. Despite these findings, Lovibond and Lovibond developed the DASS using a 4-point Likert-type scale and attached both a descriptor and number (from zero to three) to each response choice-point. In light of Cummins and Gullone's (2000) review, implementation of this response scale to measure interval data may be questionable. Consequently, the 11-point response scale anchored by zero and ten was adopted to measure depression in Australian Unity surveys from 2004 to the present with one notable exception for the purpose of investigations conducted in Study 2 in this research. The rationale for using the original 4-point scale will be explained in the method chapter in Study 2.

Differences in DASS21 response scale labels.

DASS21 response scales vary between four- and 11-point measurements formats. Additionally, clerical errors resulted in differing response scale labels within the 11-point format. These differences between the number of response points and response labels are shown in Table 11 below.

Table 11

DASS21 Response Scale Labels for 4- and 11-Point Measurement Formats

Response Scale	Survey	Response Scale Labels					
		0	1	2	3	5	10
4-point	4.2, 5.2, 8.2	did not apply to me at all	some of the time	a good part of the time	most of the time		
11-point	L1	not at all					a lot
	L2	not at all					a lot
	L3	not at all					a lot
	L4	did not apply				neutral	applied a lot
	L5	not at all					extremely
	L6	not at all					extremely

For the 4-point scale, respondents were asked to circle the number that best indicates how much the statement applied to them over the past week. For the 11-point response scale, these instructions were: 'how much do these statements apply to you over the past week?'

In terms of the scale anchors, the DASS uses severity/frequency rating scales to gauge the extent to which participants experienced negative affective states over the past week. As the temporal dimension of depression is fundamental to the clinical diagnosis of any depressive disorder, it would appear that selection of such ambiguous and differing scale anchors (10 = a lot; applied a lot; extremely) to rate the experience of negative affect over a seven-day period may impact scale reliability. Such influences may adversely affect the validity of the DASS in terms of the presumed categories of depression. The following analyses are designed to test the psychometric comparability of these various scale formats.

The number of cases in L5 that reflect primary cases was too low to be reliable (N=17), hence this survey was excluded from analyses.

Calculation of DASS depression scales.

For 4-point response choice data, depression scales were calculated as per instructions in the DASS Manual (Lovibond & Lovibond, 1995) and as described in Chapter 2. This resulted in a score range from zero to 42. Scale scores were awarded a depression severity rating (normal, mild, moderate, severe or extremely severe) according to cut-off scores derived from the normative sample data described in the DASS Manual. For the 11-point response scale data, the seven items were summed and multiplied by two; minimum score = zero, maximum score =140. Depression severity score ranges were then calculated using the formula (cut-off score/42) x (140/1). Depression severity score ranges for 4-point and 11-point response scales are presented in Table 12 below.

Table 12

DASS Depression Severity Score Ranges for 4- and 11-Point Response Formats

Severity Ratings	Score Ranges	
	4-point	11-point
Normal	0-9	0-30
Mild	10-13	30.1 – 43.3
Moderate	14-20	43.4 – 66.7
Severe	21-27	66.8 – 90
Extremely severe	28+	90.1+

Percentage of Scale Maximum (%SM)

As the DASS was originally constructed using a 4-point response scale, the key to determining comparable depression scale scores, and PWI-A scores in terms of their means and standard deviations and comparable depression severity ratings, was to convert all data to a standard form. This conversion makes it look as though all scales had been rated on a zero to 100 point scale. Stokes and Cummins (2006) have outlined the process of converting raw scores into the standard zero to 100 scale format in the PWI-A-A (English) Fourth Edition Manual (International Wellbeing Group, 2006).Values derived from this process

are called ‘percentage of scale maximum’ (%SM). Scale and item scores for depression and Personal Wellbeing for all analyses following response-scale comparisons were calculated as %SM scores according to the formula in the PWI-A Manual as follows:

$$\frac{X - k^{min}}{k^{max} - k^{min}\ x\ 100}$$

X = The score or mean to be converted

k^{min} = The minimum score possible on the scale
i.e. If a scale is score from 1 to 5, then $k^{min} = 1$
If a scale is score from -5 to +5, then $k^{min} = -5$

k^{max} = The maximum score possible on the scale
i.e. If a scale is score from 1 to 5, then $k^{max} = 5$
If a scale is score from -5 to +5, then $k^{max} = +5$

Score ranges for depressions severity ratings are presented as %SM values in Table 13 below.

Table 13

DASS Depression Severity Ratings Expressed as %SM Score Ranges

Severity Ratings	%SM	Adj. %SM range
Normal	0-21.42	0-21.42
Mild	23.81-30.95	21.43-30.95
Moderate	33.33-47.62	30.96-47.62
Severe	50 .00-64.29	47.63-64.29
Extreme	66.67 +	64.30 +

Table 13 highlights the differing range of scores within each rating of depression severity. This table also displays the discrepancies created between depression categories when converting scale scores based on the 4-point scale to %SM values. For example, the cut-off score for a rating of normal depression (21.42) does not reach the lower boundary for a rating of mild (23.81), creating a dead-space between the two ratings of 2.39 points. Similar between-ratings

discrepancies are seen across the five levels of depression. For this reason, the lower cut-off values for depression ratings were adjusted downwards to prevent omission of valid depression scale scores from analyses. Therefore, a depression rating of mild commenced at 21.43 points, moderate at 30.96 points, severe at 47.63 points, and extreme at 64.30.

Procedure

A call centre conducted telephone interviews. Interviewers read a standard set of introductory statements, instructions and questions to interviewees designed by researchers from the Australian Centre on Quality of Life.

Participation in telephone surveys was restricted to individuals eighteen years of age and over who were fluent in English. Telephone interviewers asked to speak to the person in the house who had the most recent birthday and was at least eighteen years old. Interviews proceeded on the agreement of the respondent to undertake the interview. However, they were constrained by the need to ensure that the respondent's gender and geographic location were demographically representative of the composition of the general population in Australia. Hence, gender and geographic location of survey participants were actively managed.

At the conclusion of telephone interviews, participants were invited to remain involved in the AUWI project and those who elected to do so, provided contact information to the interviewer for the purposes of mailing them a written questionnaire. To ensure participant confidentiality, a unique identifier was assigned to each participant and used to code the demographics of participants from the telephone survey, which is retained by Australian Unity. Unique identification codes were printed onto the written questionnaire. Thus, the questionnaires returned to Deakin University contained no information allowing direct identification of the names or addresses of participants. Unique identification codes of participants involved in the longitudinal follow-up were sent to Australian Unity, where codes were matched with contact information for the purposes of mailing further questionnaires. Participants were able to withdraw from the study at any time by either not returning the questionnaire or by indicating this on the returned questionnaire.

CHAPTER 5: STUDY 1 RESULTS

Data Screening and Examination of Assumptions

Prior to analysis, data were screened for missing values, and accuracy of data entry. Missing values were not replaced in the data set; rather, these were dealt with by retaining all cases and excluding cases pairwise from analyses as recommended by Pallant (2007). Several univariate outliers were detected, however, comparison of mean scores on these variables with corresponding means trimmed at the upper and lower 5% revealed that none of these outliers significantly influenced mean scores on variables germane to this study (PWI-A and DASS21-D). Consequently, univariate outliers were retained and analyses were conducted on the original data. Data were also assessed to ensure no violation of normality, linearity, homoscedasticity, multicollinearity, and singularity. For normality, Statistical Package for the Social Sciences (SPSS) descriptives identified a negative skew in the SWB scale and positive skew in the depression scale. As the skewness for both variables were within the acceptable range of -7.0 to 7.0 (Cohen & Cohen, 1983), no transformations were undertaken. Multicollinearity and singularity were assessed for both variables. As expected some correlations between depression and SWB were large ($r > .5$) (Cohen, 1988, pp. 70-81). However, as the research aim in the current study is to investigate the relationship between SWB and depression, all SWB and depression data from eight surveys were retained for analyses.

Initial Analyses

Written questionnaires in the Australian Unity Project contained two response formats for DASS21 items. Three surveys used the original 4-point response scale described previously, while 12 surveys presented participants with an 11-point end-defined response scale. In order to ascertain whether DASS21 data obtained from different surveys may be combined, examination of the psychometric comparability of the DASS21-D is required.

Internal Consistency

Cronbach's alpha coefficients for depression scales within surveys are presented in Table 14 below. All values reflect %SM scores.

Table 14
Cronbach's Alpha Coefficients for the DASS21-D Scale

Response Scale	Survey	*N*	Cronbach's α
4-point (0-3)	4.2	472	.88
	5.2	247	.89
	8.2	575	.88
	Total	1249	.88
11-point (0-10)	L1	489	.92
	L2	539	.92
	L3	634	.91
	L4	304	.93
	L6	388	.91
	Total	2354	.92

From Table 14, it is evident that the depression scale demonstrates strong internal consistency to the point of indicating item redundancy. The total range is from .88 to .93. Clearly, the response scales are not creating substantially different estimates of internal consistency.

Correlations

As detailed previously, the mean PWI-A scores are remarkably stable within Australian samples. This stability affords an opportunity to examine whether the strength of relationship between the DASS21 and a stable criterion scale varies within surveys, and may therefore differ as a function of DASS rating scales. Pearson's product moment correlations between the DASS21 and PWI-A were conducted to investigate this. The correlation coefficients are presented in Table 15 and are interpreted according to Cohen's (1988) guidelines for

determining the strength of the relationship between two variables. Specifically, correlations, r = .10 to .29 reflects a small correlation/weak relationship; r = .30 to .49 reflects a medium correlation/moderate relationship; r = .50 to 1.0 reflects a large correlation/strong relationship.

Table 15

Pearson Product-Moment Correlations between Measures of Depression and SWB

Response Scale	Survey	Scale	*r*	*N*	*M*	*SD*
4-point	4.2	1.PWI-A	-.534**	472	72.12	13.34
		2.DASS21-D			12.90	15.93
	5.2	1.PWI-A	-.612**	247	70.12	15.54
		2.DASS21-D			13.57	16.71
	8.2	1.PWI-A	-.589**	575	72.90	13.89
		2.DASS21-D			15.35	17.54
	Total	1.PWI-A	-.569**	1294	72.08	14.05
		2.DASS21-D			14.12	16.84
11-point	L1	1.PWI-A	-.588**	489	73.61	14.23
		2.DASS21-D			17.31	20.19
	L2	1.PWI-A	-.524**	539	73.16	14.86
		2.DASS21-D			19.34	20.77
	L3	1.PWI-A	-.613**	634	73.50	14.71
		2.DASS21-D			20.03	21.70
	L4	1.PWI-A	-.645**	304	74.69	14.96
		2.DASS21-D			19.24	22.91
	L6	1.PWI-A	-.626**	388	72.89	14.06
		2.DASS21-D			19.04	20.06
	Total	1.PWI-A	-.585**	2354	73.50	14.58
		2.DASS21-D			19.04	21.08

** Correlation is significant at the 0.01 level (two-tailed)

Overall, PWI-A means vary by a maximum of 4.57 percentage points across samples, only slightly higher than the variation (3.2) seen in data collected

from the sixteen cross-sectional surveys conducted to date (Cummins, Walter & Woerner, 2007).

For 4-point samples, mean DASS21-D scores vary 2.45 percentage points, and for 11-point samples by 2.75 points. Overall means for both 4-point (14.12) and 11-point (19.04) groups correspond to a DASS depression rating of normal. However, the 11-point sample means are higher overall (+4.93) than in 4-point samples and this is significant (t (3183.15) = 7.70, p = .000). The magnitude of differences in the means (mean difference = 4.92, 95% CI: 3.67 – 6.17) is small (eta squared = .016). Therefore, only 1.6% of the variance in depression is explained by response formats. Nevertheless, standard deviations related to these means are also higher (+4.26) for 11-point relative to 4-point samples. This wider range in scores suggests that the 11-point response scale demonstrates greater sensitivity to individual differences.

In line with previous empirical findings, the DASS21-D demonstrates large negative correlations with the PWI-A in all samples. The overall relationship between depression and SWB is slightly stronger in 11-point samples (-.585) than in 4-point samples (-.569), but this difference is not significant. There is also a slightly larger difference in the range of correlations in 11-point (0.121) relative to 4-point samples (.078).

Comparison of Correlation Coefficients between Two Survey Groups

To test whether correlations for two surveys are significantly different, Observed Value of Z (*Zobs*) values were calculated for all possible comparisons between surveys for 4- and 11-point response formats. Each comparison was calculated separately. These values are presented in Table 16. According to Pallant (2007), if *Zob*s is less than or equal to -1.96 or *Zobs* is greater than or equal to 1.96, correlation coefficients are statistically significantly different. If -1.96 < *Zobs* < 1.96 coefficients are not statistically significantly different.

Table 16

Zobs Values for Comparisons between Correlation Coefficients of Depression and SWB in Two Surveys

Response Scale	Surveys Compared	Z_{obs}	Significant
4-point	4.2 *v.* 5.2	-1.41891	N
	4.2 *v.* 8.2	-1.3003	N
	5.2 *v.* 8.2	0.405424	N
	Total *v.* 4.2	0.94594	N
	Total *v.* 5.2	-0.87384	N
	Total *v.* 8.2	-0.59728	N
11-point	L1 *v.* L2	1.516697	N
	L1 *v.* L3	-0.64621	N
	L1 *v.* L4	-1.2134	N
	L 1*v.* L6	-0.80612	N
	L2 *v.* L3	-2.28122	Y
	L2 *v.* L4	-2.55459	Y
	L2 *v.* L6	-2.2453	Y
	L3 *v.* L4	-0.71377	N
	L3 *v.* L6	-0.24741	N
	L4 *v.* L6	0.441907	N
	Total *v.* L1	-0.16065	N
	Total *v.* L2	1.818839	N
	Total *v.* L3	-1.0491	N
	Total *v.* L4	-1.58515	N
	Total *v.* L6	-1.14646	N

From Table 16, it is evident that there are no significant differences between the correlation coefficients for the DASS21-D and PWI-A when measuring the DASS21-D through the 4-point rating scale. For 11-point samples, the Z_{obs} values calculated for L2 and L3, L2 and L4, and L2 and L6, were -2.28, -

2.55, and -2.25 respectively. These values lie outside of the specified bounds described by Pallant (2007). Therefore, there is a statistically significant difference in the strength of the relationship between the DASS21-D and PWI-A for these surveys. As each comparison is between two different survey groups, no Bonferroni correction is required (Pallant, 2007, pp. 138-141). Depression explains significantly less variance in PWI-A for L 2 than for L3, L4, and L6.

In summary, these results indicate that the 11-point scale produces higher depression means (+4.94) that vary across a wider range of depression scores (+4.26). Nevertheless, the amount of variance in SWB explained by depression was not greater for the 11-point format.

The variation in mean scores between the 11-point and 4-point response scales are almost identical (2.45; 2.75). Thus, these two response formats demonstrate the same level of stability between surveys.

A comparison of correlation coefficients between 4-point samples revealed no significant differences. This indicates that depression explained a similar amount of variance in SWB across surveys using this format. However, there was a significant difference using the 11-point response scale in 3/15 comparisons involving L2, L3, L4 and L6.

Of these, only the DASS21-D rating scale in L3 used the same wording as L2. Scale anchors differed for the other two surveys (see Table 2). Importantly, scale anchors differed between L3, L4 and L6. However, the correlation for these three samples did not differ. Therefore, it may be concluded that the different scale labels have not differentially influenced scale reliability. Thus, data obtained from 11-point response scales will be combined in subsequent analyses.

As depression means overall were significantly higher for 11-point relative to 4-point samples, further examination is required to assess the impact response formats may have on classification of participants into categories of depression severity. As there are only three mean scores for the 4-point scales, this will be investigated at the level of individual scores.

Distribution of DASS21-D Scores

The distributions of DASS21-D scores are presented in Figures 4 to 8 below for the 4-point (zero to three) and 11-point (zero to ten) response scales.

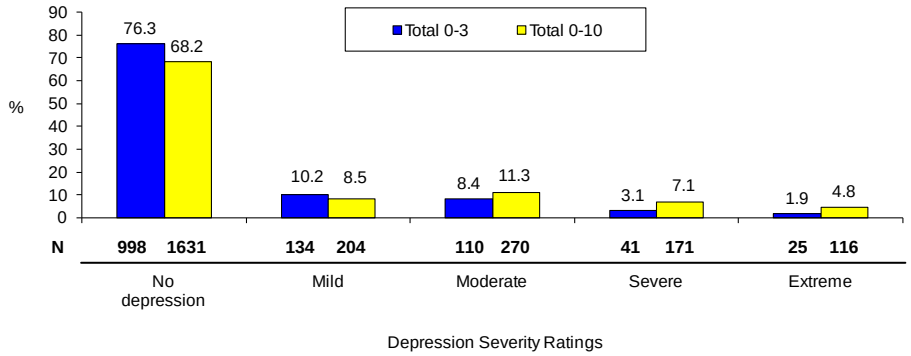


Figure 4. Distribution of DASS21-D scores across depression severity ratings

This first figure shows the combination of all PWI-A data. It is evident that the two forms of response scale have not produced the same result. More specifically:

a. The 11-point scale has identified 8.1% more people as having depression.
b. In relation to the four depression categories, the differences are less clear due to the smaller percentages involved. However, it does appear that:
c. The trend for lower identification from the 11-point scale continues into the mild category (-1.7%)
d. In the higher categories the trend is reversed, with the 11-point scale tending to identify more people in each category (+2.9%, + 4.0%, and + 2.9%)

This pattern is consistent with the 11-point scale representing a lower threshold for the diagnosis of depression classed as moderate or above. This has caused fewer people to be classed as having either no depression or mild depression. It is particularly interesting that this cross-over occurs between the mild and moderate categories. The number of points encompassed by a depression

rating of moderate increased markedly (+7.14) relative to a depression rating of mild. This factor, in combination with the downward adjustment of depression rating parameters described previously (see Table 3) influenced the classification participants' scores for the 11-point response scale into the higher (moderate and above) levels of depression. Therefore, it appears that the 11-point end-defined response scale allows for greater discrimination between individual ratings of depression as a function of the expanded response scale. Additionally, the expanded range of scores evident in depression ratings of moderate and above depression increased the potential for more participants to be classified within these levels of depression.

In sum, data obtained from the two DASS21-D response scales are not comparable. This affords an opportunity to determine further the relative advantage of the two response scales in subsequent analyses. Results will be separated by response scale depending on the nature of the question being asked.

Distribution of DASS21-D scores within specific PWI-A score ranges.

It is hypothesised that if the individual SWB score lies above 70 points using the PWI-A, the person is likely to be functioning normally and therefore, they should demonstrate a depression rating above normal. It is also hypothesised that all individuals with a SWB score of less than 50 points will evidence a depression rating above normal. In order to test these hypotheses, the distribution of DASS21-D scores for the two response scales are examined within PWI-A percentage points ranges. Hence, Figures 5 to 8 present the distribution of DASS21-D scores within the PWI-A categories of greater than 80 points, 70-80 points, 50-69 points, and PWI-A scores of less than 50 points.

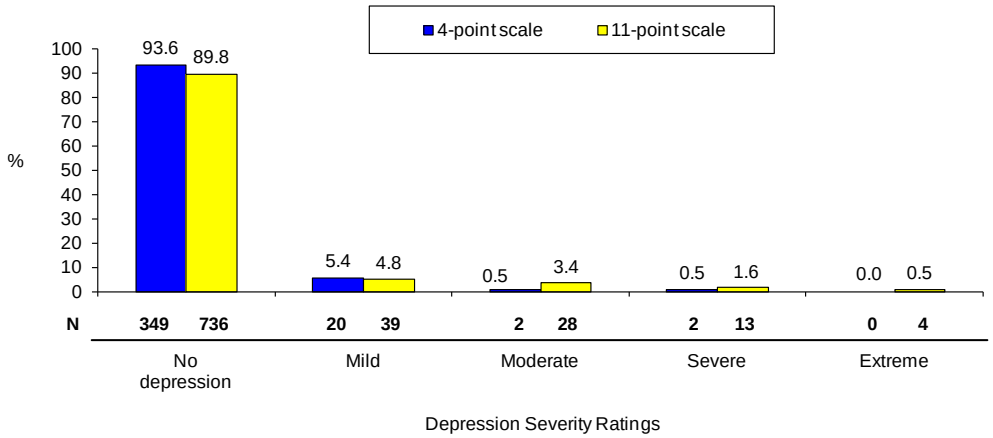


Figure 5. Distribution of DASS21-D scores within the PWI-A category >80 %SM

In the context of the theory of SWB homeostasis, it is almost inconceivable that the respondents in Figure 2 could be depressed. Yet, the PWI-A is identifying either 6.4% (4-point) or 10.3% (11-point) of these respondents as depressed.

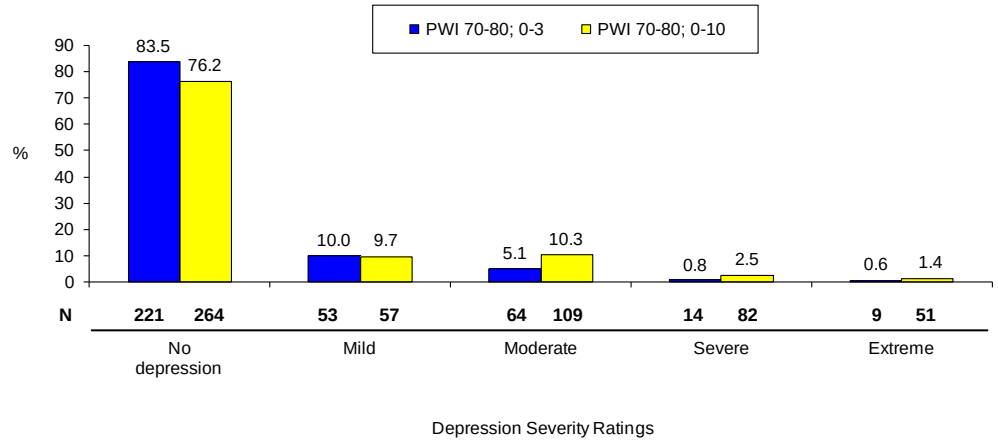


Figure 6. Distribution of DASS21-D scores within the PWI-A category 70-80%SM

It is conceivable that the 70-80 point sample might contain a few people who are depressed. For example, if someone has a set-point of 90 and a PWI-A score of 70, this 20 point difference may indicate depression. However, again the total number of people so classified by both scales (see Figure 6) far exceeds theoretical predictions within the theory of SWB homeostasis.

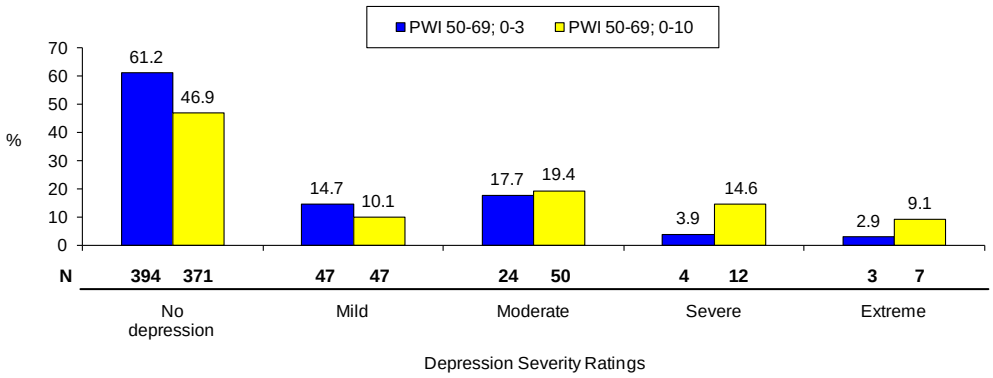


Figure 7. Distribution of DASS21-D scores within the PWI-A category 50-69%SM

As PWI-A moves into the range of 50-69 points, the proportion of people reporting a depression rating of normal drops markedly for the 4-point (-22.3) and 11-point (-29.3) response scales. The normative range for individual PWI-A scores lies between 50 and 100 and represents the 2*SD* distribution. In the context of homeostasis theory, it is possible that people with a PWI-A of 50-69 report some level of depression as a result of differing SWB set-points. However, either 38.9% (4-point) or 53.2% (11-point) of people report some level of depression and this is greater than what is theoretically expected.

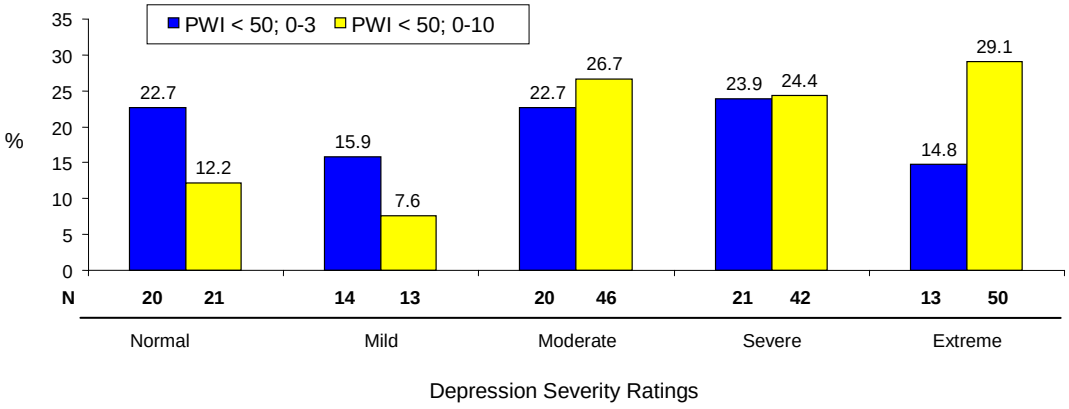


Figure 8. Distribution of DASS21-D scores within the PWI-A category <50%SM

Reviewing Figures 5 to 8 in succession suggests that the gross prediction of an inverse relationship between SWB and depression has been supported. Moreover, in accordance with expectation from homeostasis, Figure 8 demonstrates that the risk for depression increases substantially at a level of PWI-

A <50%SM. The proportion of people reporting normal, mild moderate, severe, and extreme depression within PWI-A categories is presented in Table 17.

As PWI-A decreases, there is a systematic decrease in the proportion reporting depression scores that lie within the normal limits. For example, as PWI-A scores move down from the 70-80 point range to 50-69 points, the proportion reporting normal scores falls by 29.3 percentage points. As PWI-A scores move below the theoretical threshold for individual SWB homeostasis (<50%SM), the proportion reporting normal falls a further 34.7 percentage points. However, despite reporting a PWI-A of <50%, 12.2% of people report depression scores that lie within normal limits.

Table 17

Distribution of DASS21-D Scores for 4-point and 11-Point Response Scales in PWI-A Categories

PWI-A Categories	DASS21-D Categories											
	Normal		Mild		Moderate		Severe		Extreme		% Depressed	
	4-pt	11-pt	4-pt	11-pt	4-pt	11-pt	4-pt	11-pt	4-pt	11-pt	4-pt	11-pt
PWI-A > 80%SM	93.6	89.8	5.4	4.8	0.5	3.4	0.5	1.6	0.0	0.5	6.4	10.3
PWI-A 70–80%SM	83.5	76.2	10.0	9.7	5.1	10.3	0.8	2.5	0.6	1.4	16.5	23.9
PWI-A 50–69%SM	61.2	46.9	14.7	10.1	17.7	19.4	3.9	14.6	2.9	9.1	39.2	53.2
PWI-A < 50%SM	22.7	12.2	15.9	7.6	22.7	26.7	23.9	24.4	14.8	29.1	77.3	87.8

Apparent inconsistencies with SWB homeostasis theory.

The results from these initial analyses have revealed a number of apparent inconsistencies with the theoretical predictions made from SWB homeostasis theory. Such interpretation relies on the assumption that the DASS21-D is valid in terms of the presumed categories of depression.

Cummins (2005a) asserts that only positive (PWI-A > 50%SM) is adaptive. As such, all normal SWB set-points must exist in the positive range of values (50-100%SM). If depression is conceptualised as a loss of normal levels of positive adaptive wellbeing, then the presence of depression evidences homeostatic failure and indicates a maladaptive state. Therefore, depression should not exist in combination with PWI-A of 80 points or above, and should be ubiquitous when PWI-A is less than 50 points. Further, Cummins et al. (2008, unpublished) propose that if the individual PWI-A score lies above 70 points, the homeostatic system is functioning normally, and so the person is not depressed.

Results from initial analyses dispute these predictions. Using the 11-point response scale:

a. When PWI-A is greater than 80, 10.3% of people report some level of depression and 5.05% of the sample combines at least moderate depression with a PWI-A of greater than 80 points.
b. When PWI-A is less than 50 points, 12.2% of people report a depression rating of normal.
c. When PWI-A lies between 70-80 points, 23.9% of people report some level of depression and 14.2% report moderate or above depression.

Although these inconsistencies appear to highlight substantial inconsistencies with theoretical predictions from homeostasis theory, the relationship between the PWI-A and DASS21 depression, and the constructs within the depression scale require much closer examination. Specifically, as Cummins et al. (2008, in press) has proposed that the critical evidence for the operation of homeostatic processes will be verification of a curvilinear relationship between SWB and depression scores (as found by Davern, 2004), this proposition will be investigated using a fine grain analysis of the relationship

between SWB and depression. SWB will be measured by the PWI-A, and depression will be measured by the 11-point version of the DASS21-D.

The Relationship between SWB and Depression

Sample

All data used in subsequent analyses are derived from primary cases (i.e. participants' first longitudinal measure), and are expressed as %SM scores. The sample comprises a total of 3024 participants; 57.7% males and 42.3% females aged between eighteen and 90 years *(M* = 54.05, *SD* = 15.1).

Principal Components Analysis

Lovibond and Lovibond (1995a) have determined that the DASS consists of three underlying dimensions: depression, anxiety, and stress. This study uses the depression measure only. To explore the structure of DASS depression derived from using the 11-point response format in the current general population sample, the seven items of the DASS21 were subjected to principal components analysis (PCA) using SPSS Version 12.

Prior to performing PCA, the suitability of data for factor analysis was assessed. Inspection of the correlation matrix revealed that all coefficients were above .3. The Kaiser-Meyer-Oklin value was .924, exceeding the recommended value of .6 (Kaiser, 1970, 1974) and Bartlett's Test of Sphericity (Bartlett, 1954) reached statistical significance, supporting the factorability of the correlation matrix.

Principal components analysis revealed the presence of one component with an eigenvalue exceeding one (4.70), explaining 67.15% of the variance. Inspection of the scree plot revealed a clear break after the first component. This was further supported by the results of Parallel Analysis, which showed only one component with eigenvalues exceeding the corresponding criterion values for a randomly generated data matrix of the same size (7 variables x > 2500 respondents). Item loadings on this component are presented in Table 18 below.

Table 18

DASS21-D Item Loading on PCA Component

DASS21-D Item	Component 1
Lack of interest/involvement	.867
Dysphoria	.851
Hopelessness	.849
Self-deprecation	.842
Devaluation of Life	.820
Anhedonia	.774
Inertia	.724

Item loadings from Table 18 support the current conceptualisation of clinical depression, where the two items with the highest loadings (lack of interest/involvement, dysphoria) correspond with the two requisite criteria for a diagnosis of clinical depression outlined in the DSM-IV (TR) (APA, 2001). In line with Lovibond and Lovibond's (1995) results, PCA of DASS21-D appears to support the use of one scale to summarise the seven items in this sample.

Means and Standard Deviations

Means and standard deviations for the PWI-A, DASS21-D, and DASS21-D items are presented for the total sample in Table 19.

It is interesting that there is such a difference between the mean scores of the DASS21-D items in Table 19. Such differences would appear to indicate that the items are not measuring the same thing. Conversely, the standard deviations are quite consistent, suggesting that the same distribution is moving up and down the scale. Despite these mean differences, results from Principal Components Analysis support the use of the seven depression items as one scale, as suggested by the authors of the scale (Lovibond & Lovibond, 1995).

Table 19

Means and Standard Deviations for the PWI-A, DASS21-D and DASS21-D Items

		Total Sample *N*=3,024		
		M	*SD*	Rank
Scale	PWI-A	73.40	14.54	
	DASS21-D	18.90	20.94	
Item	Anhedonia	21.34	25.25	2
	Inertia	28.68	27.44	1
	Hopelessness	16.94	26.18	5
	Dysphoria	21.15	27.55	3
	Lack of interest/involvement	18.09	25.00	4
	Self-deprecation	14.41	24.64	6
	Devaluation of Life	11.66	23.16	7

* Rank = ranking for Depression scale item within the group of depression items from 1 through to 7; where 1 = the most frequently experienced depression state, and 7 = the least frequently experienced depression state.

The PWI-A mean of 73.4 points is at the bottom of the normative Australian population range (73.43-76.43) derived from the sixteen cross-sectional surveys conducted through the AUWI Project (Cummins, Walter & Woerner, 2007).

The DASS21-D mean of 18.90 points (*SD* = 20.94) is within the normal range (0 – 21.42, see Table 13) and this is significantly higher *M* = +3.8 points, *SD* = +4.34 points; t (5936) = 7.73, $p < 0.001$, than the overall depression mean and standard deviation reported by Lovibond and Lovibond for the normative sample in the DASS Manual, *M* = 15.10%SM, *SD* = 16.60%SM. Nevertheless, as both the mean score for this sample and that of the normative sample both correspond to a DASS depression rating of normal, the depression severity ratings recommended by Lovibond and Lovibond may be used to classify cases in the current sample.

On average, participants endorsed the items reflecting (1) inertia, (2) anhedonia, and (3) dysphoria as the three most frequently experienced/most intense depression states over the past week. In the context of the manual, an equivalent depression score of between seven and eight (18.90%SM), is reported as mainly reflecting high scores on (1) inertia, (2) dysphoria, and (3) anhedonia. Thus, these results concur. This finding indicates that these three items may be the closest to the reciprocal of HPMood (Item + PWI-A = 100 points, Davern & Cummins, 2006) and as such may represent significant predictors of SWB. To investigate the relative power of the DASS21-D items to predict SWB, a standard multiple regression is performed.

Standard Multiple Regression Analysis Predicting Subjective Wellbeing by DASS21-D Items

Table 20 displays the correlations between the variables, the unstandardised regression coefficients (*B*) the standardised regression coefficients (β) the semi-partial correlations (sr^2), R^2, adjusted R^2.

The R for the regression was significantly different from zero F (7, 3016) = 243.56, p = .000. Six of seven independent variables contributed significantly to the prediction of personal wellbeing scores. These were Anhedonia (sr^2 = .006), Inertia (sr^2 = .002), Hopelessness (sr^2 = .006), Dysphoria (sr^2 = .012), Self-deprecation (sr^2 = .005), and Devaluation of life (sr^2 = .006). The seven independent variables in combination contributed another .323 in shared variability. Altogether, 36.1% (36.0% adjusted) of the variability in personal wellbeing scores was predicted by knowing scores on these seven depression variables.

Table 20

Standard Multiple Regression for Variables Predicting Personal Wellbeing by DASS21-D Scale Items (N=3,024)

Variables	PWI-A	1	2	3	4	5	6	*B*	β	sr^2
1. Anhedonia	-.47							-.065**	-.112	0.006
2. Inertia	-.40	.54						-.030*	-.057	0.002
3. Hopelessness	-.52	.61	.52					-.071**	-.128	0.006
4. Dysphoria	-.53	.61	.55	.69				-.095**	-.179	0.012
5. Lack interest/involv.	-.48	.40	.62	.68	.70			-.000	.000	0.000
6. Self-deprecation	-.51	.62	.51	.66	.67	.68		-.072**	-.122	0.005
7. Devaluation of life	-.50	.56	.46	.69	.64	.65	.73	-.079**	-.125	0.006
R^2 = .361							Total explained unique variance		= .037	
Adjusted R^2 =.360							Total explained shared variance		= .323	

*p<.005; **p<.001

Although the correlation between personal wellbeing and lack of interest/involvement was -.48, lack of interest/involvement did not contribute significantly to the regression. Post-hoc evaluation of the correlation between personal wellbeing and lack of interest/involvement revealed that it was significantly different from zero F (7, 3016) = 87.06, $p < 0.01$. Apparently, the relationship between personal wellbeing and lack of interest/involvement is mediated by the relationships between anhedonia, inertia, hopelessness, dysphoria, self-deprecation, devaluation of life, and personal wellbeing. This regression demonstrates that all DASS21-D items, with the exception of lack of interest/involvement are significant predictors of personal wellbeing. It is particularly interesting that lack of interest/involvement was not a significant predictor, as this construct meets one of the two criteria that must be met for the diagnosis of any clinical depressive disorder, the other being lowered mood (anhedonia).

Of the DASS21-D items, dysphoria has the highest correlation with PWI-A (-.53) and contributes the greatest amount of unique variance in the regression (sr^2 = .012). This suggests that of the seven items, dysphoria, a measure of abstract negative affect, more closely resembles the reciprocal of HPMood (abstract pleasant affect). If this proposition is correct then the dysphoria item should demonstrate a strong inverse linear relationship with the PWI-A.

Examination of PWI-A means in the context of DASS21-D item means was performed to investigate the nature of the relationship between SWB and the states of DASS21 depression.

Means and Standard Deviations across Depression Severity Ratings

Means and standard deviations for the PWI-A, depression scale and depression items are presented within depression severity ratings in Table 21. These mean values are plotted in Figure 9. Table 22 contains percentage point changes in mean values between consecutive ratings of depression severity.

Table 21

Means and Standard Deviations for PWI-A, DASS21-D Items within DASS Depression Severity Ratings

	DASS Depression Severity Ratings														
	Normal *N*=2067			Mild *N*=255			Moderate *N*=345			Severe *N*=211			Extreme *N*=146		
	M	*SD*	Rk	*M*	*SD*	Rk	*M*	*SD*	Rk	*M*	*SD*	Rk	*M*	*SD*	Rk
PWI-A	78.33	10.83		70.81	12.42		63.40	13.90		59.74	15.22		50.18	17.00	
DASS21-D.	6.86	6.40		26.16	2.46		38.95	4.94		55.62	4.61		76.16	8.78	
An.	9.60	15.93	2	34.35	21.34	2	43.30	19.65	3	51.75	18.16	6	69.32	20.09	7
In.	16.92	19.75	1	42.00	23.25	1	50.81	22.50	1	60.04	19.46	2	74.32	21.49	5
Hop.	3.91	10.07	5	24.90	22.25	5	38.23	23.09	4	56.73	19.77	3	79.73	16.19	2
Dys.	7.41	13.01	3	28.63	20.93	3	46.75	22.77	2	62.09	19.58	1	83.08	14.41	1
Lack int	5.38	10.46	4	26.47	19.56	4	40.43	20.41	5	55.83	17.03	4	76.00	18.25	4
Self-dep	3.19	8.23	6	15.76	17.79	6	29.22	23.81	6	54.55	21.96	5	77.88	19.41	3
Deval.life	1.63	6.55	7	10.98	14.93	7	23.88	23.97	7	48.34	23.04	7	72.88	23.04	6

Rk = ranking for Depression scale item within the group of depression items from 1 through to 7; where 1 = the most frequently experienced depression state, and 7 = the least frequently experienced depression state.

Table 22

Percentage Point Change between PWI-A, DASS21-D Scale and Depression Item Means in Consecutive DASS Depression Severity Ratings

	Mean Change			
	Normal-Mild	Mild-Moderate	Moderate-Severe	Severe-Extreme
PWI-A	-7.52	-7.41	-3.66	-9.56
DASS21-D	19.3	12.79	16.67	20.54
Anhedonia	24.75	8.95	8.45	17.57
Inertia	25.08	8.81	9.23	14.28
Hopelessness	20.99	13.33	18.5	23.00
Dysphoria	21.22	18.12	15.34	20.99
Lack of interest/involvement	21.09	13.96	15.40	20.17
Self-deprecation	12.57	13.46	25.33	23.33
Devaluation of life	9.35	12.90	24.46	24.54

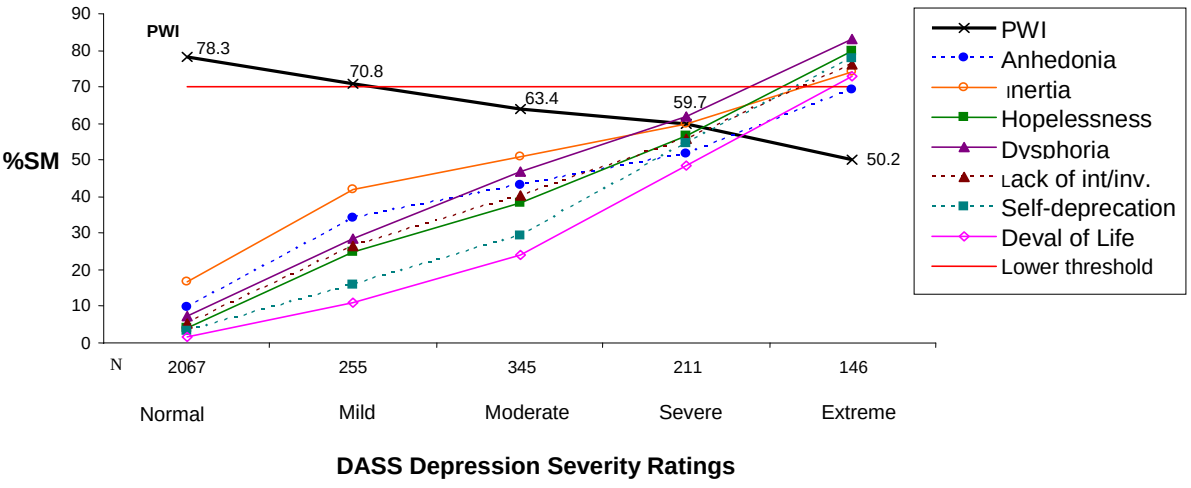


Figure 9. Mean values for PWI-A and DASS21-D items in DASS depression severity ratings

It is evident that there is a strong, inverse, linear relationship between each of the DASS21-D items and the PWI-A scores across the depression severity

ratings, however, there are differences between the items. These differences are highlighted in Figure 8 and Table 23.

From Table 21, mean depression scores increase by 69.30 points across the depression rating of normal to extreme, whereas the PWI-A means decrease by 28.15 percentage points. This is suggestive of PWI-A resisting change and may provide some evidence for homeostasis.

Across this same range, the incremental decreases in PWI-A means become progressively smaller up to a rating of severe depression (-7.52, -7.41, -3.66), before falling markedly at a depression rating of extreme (-9.56). This pattern is consistent with the plateau, or at least a slowing of the drop. This pattern of change is reciprocal to that demonstrated by the items reflecting anhedonia and dysphoria. Specifically, the rate of change slows over two points for anhedonia (see Table 22) at mild-moderate (-8.95), and then slows slightly further at moderate-severe (-8.45). This is the same pattern seen for dysphoria (mild-mod-18.25, mod-sev-15.34). Inspection of Figure 9 reveals differences in the responsiveness of anhedonia and dysphoria across increasing depression scores, despite similar trends in mean change highlighted in Table 22. Whereas the mean value for the dysphoria items continues to increase as the full-scale depression scores increases, the anhedonia item shows a slower rate of increase beyond a depression severity rating of mild. A slight flattening of the slope from a rating of mild depression evidences this. Additional support for the differences in responsiveness is found in the ranking changes for these items in Table 22.

Notably, the PWI-A mean falls below the normative population range threshold for SWB at a depression rating of moderate.

However, comparing rates of mean change across depression severity ratings may be confusing and misleading. Depression severity ratings comprise differing ranges of depression scale points. These are presented in Table 23 below.

Table 23

Number of Scale Points within DASS Depression Severity Ratings

Depression Severity Ratings	Scale points within Depression Ratings	
	4-point	%SM
Normal	9	21.42
Mild	4	10.52
Moderate	7	17.66
Severe	7	17.66
Extremely severe	15	36.7

The differing number of scale points within levels of depression displayed in Tale 23 suggests that the rate of change in PWI-A and depression item means may be an artefact of the categories of depression severity. Therefore, the use of only these five levels of depression may have obscured any plateau effect. The following analysis will group depression scale scores according to a larger number of intervals of equivalent range not based on the depression classification criteria.

PWI-A and Depression Means across Increments of DASS Depression Scores

Table 24 is based on DASS depression score increments of 7.14 points. This increment size is the smallest range that allows an N >20 per group for most groups. The final two increments (DASS-D scores 36.1-39.0, and 39.1-42.0) are not considered in the following analysis, as the number of cases per groups is too small for them to be reliable.

Table 24

Means and Standard Deviations for PWI-A and DASS21-D across Increments of Depression Scale Scores

Depression Rating	Depression Score	Scale	N	% of Total	Min.	Max.	M	*SD*	Inc. Change
Normal 0-9	0	PWI-A	484	16.01	35.71	98.57	82.80	10.09	
		DASS21-D	484		.00	.00	.0000	.00000	
	0.1-3.0	PWI-A	785	25.96	11.43	98.57	79.30	10.01	-3.49
		DASS21-D	785		1.43	7.14	3.93	2.01	3.93
	3.1-6.0	PWI-A	497	16.44	17.14	97.14	75.52	10.72	-3.78
		DASS21-D	497		8.57	14.29	11.26	2.02	7.32
	6.1-9.0	PWI-A	301	9.95	14.29	95.71	73.26	10.87	-2.27
		DASS21-D	301		15.71	21.43	18.27	1.925	7.02
Mild 10-13	9.1-12.0	PWI-A	218	7.21	34.29	94.29	70.48	12.65	-2.78
		DASS21-D	218		22.86	28.57	25.50	2.03	7.23
Moderate 14-20	12.1-15.0	PWI-A	161	5.32	25.71	88.57	67.73	13.23	-2.75
		DASS21-D	161		30.00	35.71	32.72	2.10	7.22
	15.1-18.0	PWI-A	142	4.70	32.86	95.71	63.21	13.37	-4.52
		DASS21-D	142		37.14	42.86	39.90	1.96	7.18
	18.1-21.0	PWI-A	115	3.80	17.14	94.29	61.29	14.30	-1.92
		DASS21-D	115		44.29	50.00	46.91	2.04	7.01
Severe 21-27	21.1-24.0	PWI-A	106	3.51	18.57	97.14	60.42	14.87	-0.87
		DASS21-D	106		51.43	57.14	54.02	1.89	7.11
	24.1-27.0	PWI-A	69	2.28	14.29	91.43	58.47	16.97	-1.95
		DASS21-D	69		58.57	64.29	61.28	1.82	7.27
Extreme 28+	27.1-30.0	PWI-A	62	2.05	4.29	81.43	50.81	14.43	-7.66
		DASS21-D	62		65.71	71.43	68.48	2.07	7.20
	30.1-33.0	PWI-A	40	1.32	18.57	91.43	47.70	16.00	6.62
		DASS21-D	40		72.86	78.57	75.54	1.95	7.06
	33.1-36.0	PWI-A	23	0.76	12.86	74.29	44.60	14.80	-12.83
		DASS21-D	23		80.00	85.71	83.04	2.25	7.51
	36.1-39.0	PWI-A	15	0.50	12.86	75.71	42.00	18.40	-2.60
		DASS21-D	15		87.14	92.86	90.48	1.76	7.43
	39.1-42.0	PWI-A	6	0.20	8.57	81.43	37.14	30.81	-4.86
		DASS21-D	6		94.29	100.00	97.62	1.95	7.14

From Table 24, the following observations can be made:

1. PWI-A falls 38.20 percentage points between a depression score of zero (normal) and 36.0 (extreme). Somewhat surprisingly, even at depression scores of 21.1 through 27.0 (severe), PWI-A means remain within the positive range of values (60.42, 58.47). Across the depression score increment of 30.1-33.0 (extreme), the level of

SWB remains positive on average (57.43). This is almost inconceivable in the context of SWB homeostasis theory.

2. Depression scores increase a total 83.04 percentage points from a depression rating of zero to 33.1-36.0. This represents more than double the change for the PWI-A. The incremental change between depression means remains steady due to the constraints of using depression scale scores as the grouping criterion.
3. The PWI-A means fall 12.32 percentage points from a depression score of zero up to a score of 9.1-12 (mild) where they approximate the lower homeostatic threshold for group means (70.48 points). Over this range, the change in PWI-A means between increments of depression varies no more than 1.01 percentage point, evidencing a relatively stable rate at which PWI-A percentage points are lost. However, at the crossover between depression scores of 9.1-12 and 12.1-15 (moderate) depression, PWI-A means fall below the lower threshold from 70.48 to 67.73 points (-2.75) yet the incremental change between groups of depression scores remains steady. Over the next increment (15.1-18), PWI-A means fall a further 4.52 percentage points (M = 63.21) and this represents an increase in the rate at which levels of SWB are falling. Over the following three increments, (18.1-21, 21.1-24, 24.1-27) the rate at which PWI-A means continue to fall slows markedly (-1.92, -0.87, -1.95). These percentage point changes represent the three smallest losses in PWI-A means over all depression score increments and correspond to the crossover between depression ratings of moderate and severe.

From these data, it appears slightly possible that a plateau effect, where PWI-A means hold relatively steady despite increasing depression scores, is present. However, the standard deviations associated with PWI-A means do not appear to be decreasing and therefore do not represent a narrowing of the overall range of values across these score increments commensurate with such a plateau Nevertheless, it is possible that if the means are continuing to fall, but at a slower rate, the *SD* may not be expected to change, or to change only slightly. Given that

the amount of information contained within Table 24 may potentially obscure the relationship between the PWI-A and DASS21-D means, these means are plotted across depression score increments in Figure 10 below.

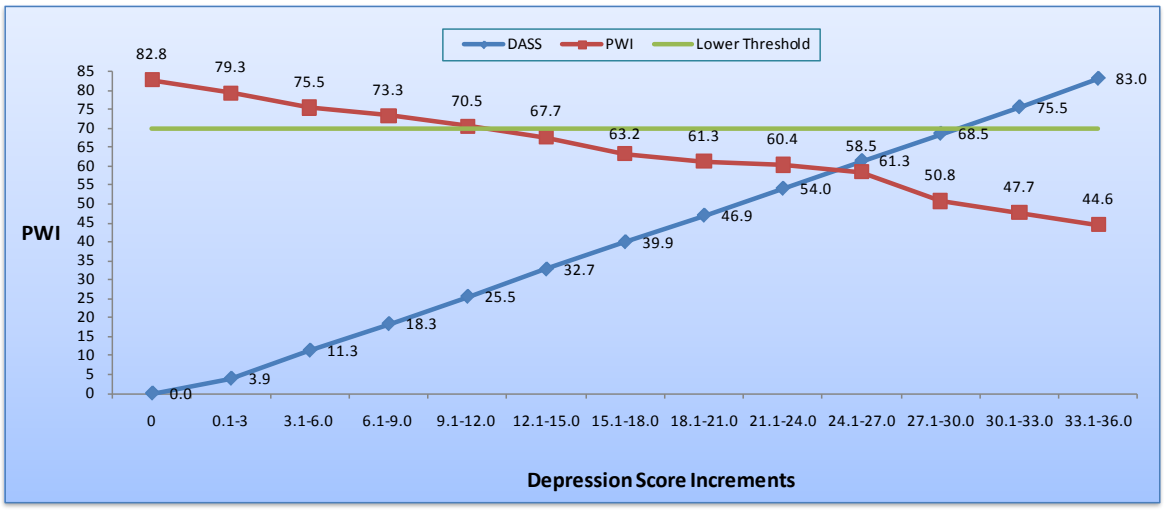


Figure 10. PWI-A and DASS21-D means across increments of depression scores

Figure 10 reveals a gradual and variable decrease in PWI-A means across depression score increments. The rate at which PWI-A means decrease appears to slow at a depression score of 15.1-18 (high moderate). From this increment, PWI-A appears to hold relatively steady from scores of 18.1 through to 27, before falling sharply (-7.66) to approximate the theoretical absolute threshold for homeostasis (50.8 points; extreme). Over the following two increments PWI-A means fall steadily (-3.1, -3.1) to lie below the theoretical absolute threshold for homeostasis (47.7, 44.6).

Taken together, these results support Davern's (2003) findings in so much as PWI-A means in this sample fall below the lower homeostatic threshold for group means at a depression rating of moderate using the DASS21. However, although there appears to be a relatively marked reduction in PWI-A means (-4.52) over the following increment (15.1-18; moderate), PWI-A means do not continue to fall sharply thereafter as proposed by Cummins et al. (2008). Therefore, this decrease is not indicative of homeostatic failure at this level. Rather, in this sample, over the depression scores of eighteen through 27, PWI-A means lost the smallest number of percentage points for any depression increments. This evidences a reduced rate at which levels of SWB were declining

despite the increasing presence of depressive symptoms and points to a possible plateau effect. Nevertheless, a definite plateau in PWI-A means remains uncertain and requires further testing.

Verification of a plateau effect may be undertaken in the following manner. If homeostasis fails at a depression rating of moderate as Cummins et al. (2008, unpublished) suggest, then it may be reasonable to expect that PWI-A means will differ significantly from each other at the point where the lower threshold for group means is breached. That is, the difference between the PWI-A mean of 70.48%SM will differ significantly from the PWI-A mean of 67.73%SM. Further, that there will be no significant differences in PWI-A means between depression increments up to and including a depression rating of mild (0.1-3 to 9.1-12), as SWB is being defended by homeostatic processes. A one-way between groups analysis of variance is required to test these propositions and allows for the testing of a possible plateau effect.

If a plateau effect is present over the three increments (15.1-18 and 18.1-21, 18.1-21 and 21.1-24, 21.1-24 and 24.1-27) that demonstrate the smallest incremental changes in PWI-A means, then PWI-A means will differ in the lead up to, and after this group of score increments, and will not differ significantly between these score increments. That is, there will be a significant difference between the PWI-A mean 67.73 and 63.21, and 58.47 and 50.81, and there will be no significant differences between the PWI-A means of 63.21, 61.29, 60.42, and 58.47.

Analysis of Variance of PWI-A Means according to DASS21 Depression Score Increments

A one-way between groups analysis of variance was conducted to explore the impact of DASS21 depression scores on levels of personal wellbeing. The first (0) and final two score increments (36.1-39.0, 39.1-42.0) were not included in this analysis for reasons described earlier.

There was a statistically significant difference at the $p<0.001$ level in Personal Wellbeing means for the twelve depression score groups according to the Welch statistic [Welch (11,299.05) = 78.90, p = .000]. The Welch statistic is

preferable to the F statistic when the assumption of equal variances does not hold. Due to the large volume of comparisons generated by this test, only comparisons of PWI-A means between consecutive depression score increments are reported as these mean differences are germane to investigating Cummins et al.'s (2008) propositions and examining the possibility of a plateau effect at the cross-over between depression ratings of moderate and severe.

Post-hoc comparisons using the Dunnett T3 indicated that the PWI-A means for those with a depression score of 0.1-3.0 were significantly different to those with a depression score of 3.1-6.0 (p = .000). There were no other significant differences in PWI-A means between consecutive increments of depression scores.

These results do not support Davern's (2003, unpublished) findings and Cummins' theoretical proposition of homeostatic failure at the crossover between depression ratings of mild and moderate. There were no significant differences between PWI-A means over the three increments (18.1-21.0, 21.1-24.0, 24.1-27.0) that suggested the possibility of a plateau effect. However, the absence of significant differences between consecutive PWI-A means before and after these increments makes it highly unlikely that SWB levels evidence a distinct plateau that represents a statistically significant change in its relationship to DASS21 depression. Further examination of this plateau effect is conducted in the analyses described below.

Comparison of Slope

The following comparisons were undertaken using a Slope Comparison Calculator (SCC) that was designed by Associate Professor Mark Stokes (2007) in line with statistical principles outlined in Zar's (1999) text. Data point co-ordinates plotted on the *x*- and *y*-axis for two slopes within one variable or the slopes of two variables are entered into the calculator spreadsheet and a t-test is used to assess whether the two slopes significantly differ. From Figure 11, it is evident that plotting PWI-A means across increasing DASS21-D scores reveals two different slopes. The data for these two slopes are presented following Figure 11 in Table 25.

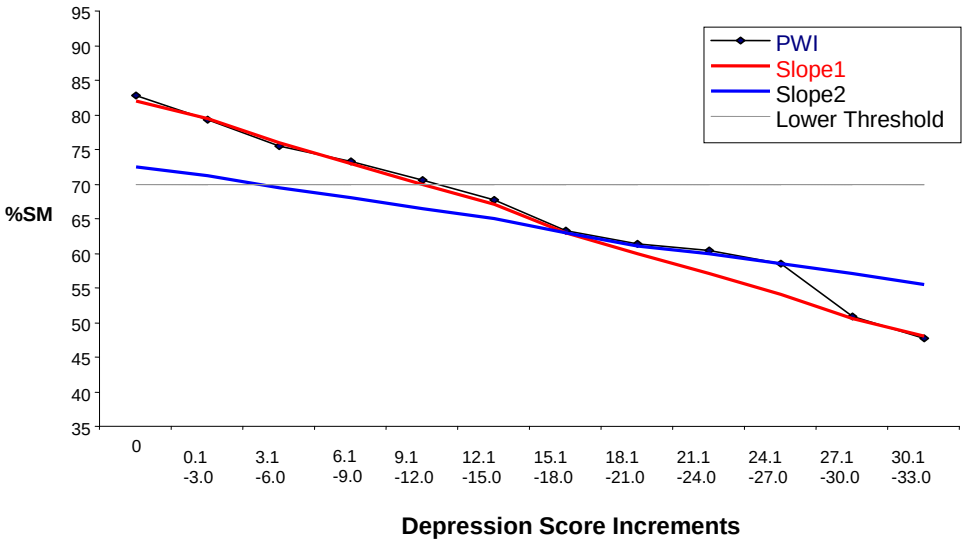


Figure 11. Lines of best fit through PWI-A means across DASS21-D score increments

These two apparently different trajectories (slopes) for mean PWI-A values are highlighted in Figure 11 and the *x* and *y* co-ordinates for the two slopes are detailed in Table 26.

Table 25

Data Table for Figure 11

x-axis	DASS21-D score	0	1.55	4.55	7.55	10.55	13.55	16.55	19.55	22.55	25.55	28.55	31.55
y-axis	PWI-A	82.8	79.3	75.5	73.3	70.5	67.7	63.2	61.3	60.4	58.5	50.8	47.7
	Slope1	82.0	79.5	76.0	73.0	70.0	67.0	63.0	60.0	57.0	54.0	50.6	48.0
	Slope2	72.5	71.2	69.5	68	66.5	65.0	63.0	61.0	60.0	58.5	57	55.5

Three sets of *x*- and *y*-axis co-ordinates were selected for comparison of slope, and these values are shaded in Table 25. The PWI-A and Slope 1 means represent the *y*-axis co-ordinates for the two slopes. The corresponding DASS21-D score increments were each averaged to form the *x*-axis co-ordinates for the two slopes. These *x*- and *y*-axis values were entered into the SCC and calculations revealed that there was no significant difference between the slopes within PWI-A means across depression score increments; t (2) = 0.40, p = .73 (two-tailed). The magnitude of the differences in means (mean difference = .53) was very small (eta squared = .0009). A possible problem with this procedure is the N=3 points for the comparison line, which gives very small *df*, thereby decreasing the sensitivity of the statistical test.

Notwithstanding this issue, from these analyses, it appears unlikely that a curvilinear relationship exists between SWB and DASS21 depression. In these terms, this finding may indicate that depression, as measured by the DASS21, approximates the reciprocal of HPMood, and as such may be viewed as an outcome of homeostatic failure rather than a challenge to SWB homeostasis. To test this interpretation, the relationship between SWB and the constructs comprising the depression scale were more closely examined in the following analyses.

PWI-A and DASS21-D Item Means and Standard Deviations

Inspection of PWI-A and DASS21-D items means within depression severity ratings were obfuscated by the different number of scale points within rating categories. To overcome this problem, means and standard deviations are presented within depression score increments of equivalent range in Table 26 and Figure 12 below.

Table 26

Means and Standard Deviations for the PWI-A and DASS21-D Items within Depression Score Increments

Depression Severity	DASS21-D Score Increments	x-axis	N	%		PWI-A	Anh.	Ine.	Hop.	Dys.	Lack int/inv.	Self-dep.	Dev. of Life
Normal	0		484	16.0	M	82.80	0	0	0	0	0	0	0
0-9					*SD*	10.09	0	0	0	0	0	0	0
	0.1-3.0	1.55	785	26.0	M	79.30	5.30	12.62	1.54	4.22	2.24	1.15	.46
					SD	10.01	9.57	11.50	5.33	7.64	5.23	4.07	2.73
	3.1-6.0	4.55	497	16.4	M	75.52	16.30	27.67	6.12	11.73	8.77	5.49	2.72
					SD	10.72	17.44	21.09	10.66	13.70	11.27	8.86	7.10
	6.1-9.0	7.55	301	10.0	M	73.26	25.05	37.57	12.76	20.50	16.68	9.83	5.52
					SD	10.87	22.25	21.83	17.65	19.63	15.90	14.62	12.94
Mild	9.1-12.0	9.55	218	7.2	M	70.48	34.08	41.06	24.63	28.07	25.64	14.17	10.88
10-13					*SD*	12.65	21.58	23.40	22.52	21.36	19.34	16.61	15.41
Moderate	12.1-15.0	13.55	161	5.3	M	67.73	38.26	46.21	32.61	39.69	33.42	23.66	15.22
14-20					*SD*	13.22	19.73	21.56	22.51	22.76	20.13	22.49	19.11
	15.1-18.0	16.55	142	4.7	M	63.21	45.42	51.83	36.69	46.90	41.20	30.21	27.04
					SD	13.37	19.15	22.24	22.24	21.51	17.40	23.108	24.72
	18.1-21.0	19.55	115	3.8	M	61.30	48.35	55.57	47.82	54.70	49.57	39.48	32.87
					SD	14.30	18.21	21.77	22.47	21.00	21.33	23.54	23.69
Severe	21.1-24.0	22.55	106	3.5	M	60.42	51.51	59.91	53.96	59.81	54.15	52.74	46.04
21-27					*SD*	14.87	18.14	19.05	19.01	19.95	15.73	21.85	22.66
	24.1-27.0	25.55	69	2.3	M	58.47	51.59	64.06	64.63	68.55	61.16	62.17	56.81
					SD	16.97	20.19	19.50	17.87	18.01	17.02	20.57	22.33
Extreme	27.1-30.0	28.55	62	2.1	M	50.81	59.03	67.10	72.10	78.23	67.42	69.52	65.97
28+					*SD*	14.43	18.79	22.93	14.04	14.20	18.10	19.28	23.71

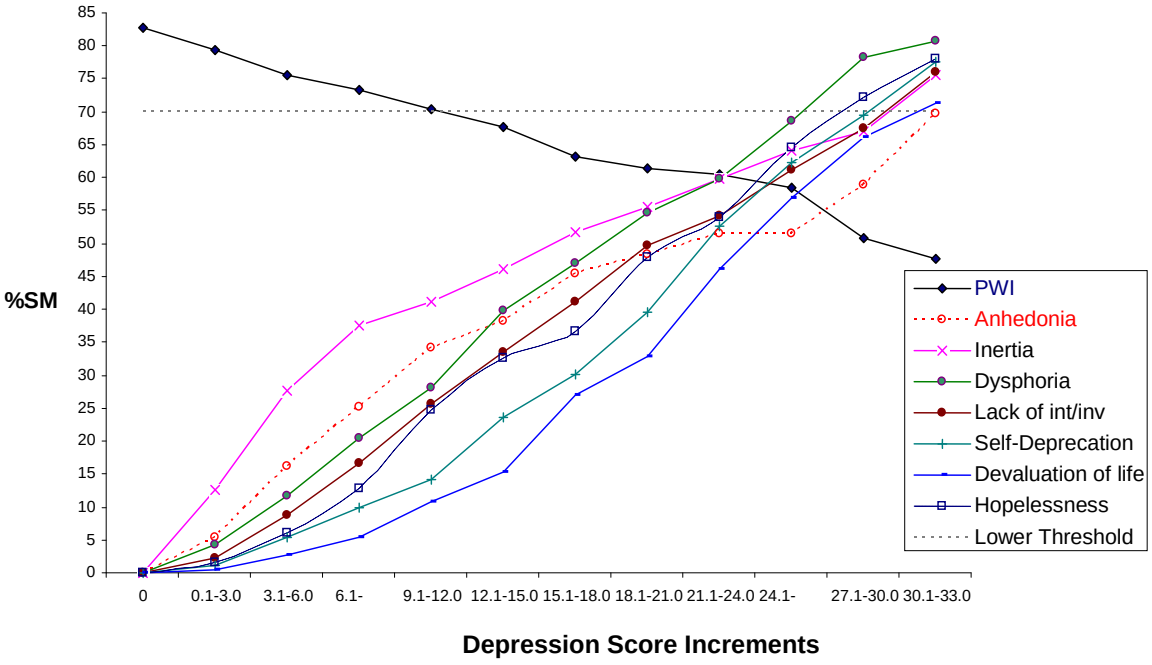


Figure 12. PWI-A and DASS21-D item means within depression scores increments

Figure 12 depicts a strong inverse linear relationship between all of the DASS21-D items and the PWI-A. However, it is difficult to determine which of the DASS21-D items follows the same pattern of change as the PWI-A. To investigate this, slopes of the PWI-A and DASS21-D items were compared using the SCC described previously. Seven comparisons were performed between the PWI-A and each of the DASS21-D items. Each depression score increment was averaged (see Table 12) to form the *x*-axis co-ordinates for the slopes, and corresponding PWI-A and DASS21-D item means comprised the *y*-axis co-ordinates for the respective slopes. No statistical correction was required as a separate analysis was performed for each comparison. Results for slope comparisons are presented in Table 27.

Table 27

Comparison between the Slope for PWI-A and the Slopes for the DASS21-D Items

Slope 1	Slope 2	*p*- value (2-tailed)
PWI-A	Anhedonia	.07
PWI-A	Inertia	.16
PWI-A	Hopelessness	.07
PWI-A	Dysphoria	.13
PWI-A	Lack of interest/involvement	.07
PWI-A	Self-deprecation	.04*
PWI-A	Devaluation of life	.02*

From Table 27, there was no significant difference between the PWI-A slope and the slope of anhedonia, inertia, hopelessness, dysphoria or lack of interest/involvement. There was a significant difference between the PWI-A slope and the slope of self-deprecation: $t(16) = 2.2$, $p = .04$ (two-tailed). There was also a significant difference between the PWI-A slope and the slope of devaluation of life: $t(16) = 2.6$, $p = .02$ (two-tailed). Additionally, the means for these two items (self-deprecation; devaluation of life) are the least like the PWI-A (see Table 26). This is consistent with an idea that they are more cognitive, perhaps, and so less driven by HPMood, and so more like causal (challenge) than indicator variables. This may account for the visual emergence of a plateau. In these terms, failure to statistically verify the presence of a significant plateau may be because the other five of the seven DASS21-D items essentially measure the reciprocal of HPMood (i.e. a lack of positive affect), and therefore function as outcome variables, just like SWB. Clearly these results require replication.

Results Summary

To briefly summarise these results:

- Principal Components Analysis of depression supports the use of one component to summarise the seven depression items.

- There is a strong, inverse linear relationship between each of the DASS21 depression items and the PWI-A scores across ratings of depression severity.
- Accurate comparison of mean change in PWI-A and DASS21 depression cannot be made across ratings of depression severity due to the differing number of depression scale points within severity rating groups.
- The mean PWI-A for the total sample ($\underline{M}$ = 73.40, *SD* = 14.54) lies within the parameters for the normative population range for SWB of 70-80 points, and represents the bottom of the normative range for Australian samples.
- PWI-A means fall below the lower normative threshold (<70) at a depression severity rating of moderate using the DASS21-D (63.40). This result corroborates Davern's (2003, unpublished) findings.
- At a depression severity rating of extreme, the PWI-A mean remains in the positive wellbeing range (50.18).
- DASS21-D means increase 83.04 percentage points from a depression score of zero (normal) to 36.0 (extreme). This represents more than double the change for PWI-A across the same range of scores (38.20 points).
- The mean and standard deviation for DASS21 depression ($\underline{M}$ = 18.90, *SD* = 20.94) indicate that participants report experiencing the states of depression at a frequency considered normal for a general population sample.
- The DASS21-D mean and standard deviation for this sample are higher (*M* = +3 points, *SD* = +4.34 points) than those reported by Lovibond and Lovibond for the normative sample in the DASS Manual (*M* = 15.10%SM, *SD* = 16.60%SM).
- DASS21-D item responsiveness differs between items and across depression score increments.
- The DASS21-D items reflecting inertia, anhedonia, and dysphoria, demonstrate the highest mean values in the context of a normal mean depression scale score. This indicates that these three items are more

responsive at low levels of depression than other depression scale items.

- Standard multiple regression predicting personal wellbeing by DASS21-D items demonstrated that 36.1% of the variance in the PWI-A may be predicted by knowing the scores on six of the seven items of the DASS21 depression scale. The item reflecting lack of interest/involvement did not significantly contribute to the regression. Post-hoc evaluation of the correlation between personal wellbeing and lack of interest/involvement revealed that it was significantly different from zero. Apparently, the relationship between personal wellbeing and lack of interest/involvement is mediated by the relationships between the other six DASS21-D items and personal wellbeing.
- A plateau effect appears to emerge across the depression score increments (18.1-21.0, 21.1-24.0, 24.1-27.0) where PWI-A means seem to hold relatively steady despite increasing depression scores.
- One-way analysis of variance of PWI-A means across depression score increments with post-hoc comparisons using the Dunnett T3 revealed a significant difference between means across the depression score increments of 0.1-3.0 compared to 3.1-6.0. There were no other significant differences between PWI-A means across any other consecutive depression score increments. This finding disputes the propositions of a homeostatic failure at a DASS21 depression rating of moderate, and the existence of a plateau effect.
- Inspection of PWI-A means across DASS21-D score increments revealed two apparently different slopes. Comparison of these slope revealed that they did not significantly differ. This finding corroborates results from the analysis of variance.
- Examination of PWI-A and DASS21-D items means across depression score increments confirmed a strong, inverse linear relationship between the PWI-A and all DASS21-D items.
- In the context of depression score increments, the slope for PWI-A means was compared with each of the slopes for DASS21-D item means. These comparisons revealed that the slope for personal

wellbeing was similar to the slopes of five of the DASS21-D items, anhedonia, inertia, hopelessness, and dysphoria, and differed significantly from the slopes for self-deprecation and devaluation of life.

CHAPTER: 6 STUDY 1 DISCUSSION

The major aim of this study was to investigate the propositions made within the theory of SWB homeostasis (Cummins, 2009, in press) regarding the relationship between SWB and depression. Specifically, this study sought to test the theoretical predictions and diagnostic approximations regarding the relationship between SWB and depression. SWB was measured by the PWI-A-A (International Wellbeing Group, 2006) and depression was measured by the DASS21-D (Lovibond & Lovibond, 1995a).

To address these research aims, a number of hypotheses were proposed regarding the relationship between SWB and depression. These hypotheses were that:

1. Consistent with previous empirical research and SWB homeostasis theory, depression will demonstrate moderate negative correlations with SWB.
2. If the individual SWB score lies above 70 points using the PWI-A, the person is likely to be functioning normally, and therefore it is highly unlikely that they will demonstrate a rating on the DASS21-D consistent with depression.
3. All individuals with a SWB score of less than 50 points will demonstrate a rating on the DASS21-D consistent with depression.
4. In confirmation of the operation of homeostatic processes to maintain normal levels of SWB, SWB scores will demonstrate a curvilinear relationship with increasing depression scores.

Hypothesis One

In line with previous empirical research and SWB homeostasis theory, SWB demonstrate a moderate to strong inverse relationship with depression.

Results supported the hypothesis. In line with previous empirical findings (Cheung & Bagley, 1998; Davern, 2004; Goldberg & Harrow, 2005; Hansson, 2002; Heady, Kelly & Wearing, 1993; Hong & Giannakopoulos, 1994; Koivumaa-Honkanen et al., 2001, 2004; Lewis et al., 1999; Simpson et al., 1996;

Reid et al., 2003), there was a strong inverse relationship between SWB as measured by the PWI-A, and depression as measured by the DASS21-D using 4-point response formats in three surveys (-.57) and 11-point response formats in six surveys (-.59). While there was a trend towards a slightly stronger association in 11-point samples, the difference between the response formats was not significant. These findings are consistent with the theoretical proposition that depression represents the failure of the homeostatic system to maintain SWB, and the subsequent loss of the normal positive sense of HPMood (Cummins et al., 2009). In these terms, it may be reasonable to expect that an individual reporting a PWI-A that lies within the population normative range of 70 to 80 points (Cummins 1995, 1998, 2003) would be unlikely to report a DASS21-D score consistent with depression. Alternatively, PWI-A scores that lie within the negative (dissatisfied; < 50 points) sector of the dissatisfaction-satisfaction continuum, should correspond to a DASS21-D rating of depression. These two predictions made within SWB homeostasis theory, are tested in the next two hypotheses, and will be discussed together.

Hypothesis Two and Hypothesis Three

If the individual SWB score lies above 70 points using the PWI-A, the person is likely to be functioning normally and therefore, it is highly unlikely that they will demonstrate a rating on the DASS21-D consistent with depression.

It is highly likely that individuals with a PWI-A score of less than 50 points will demonstrate a rating on the DAS21-D consistent with depression.

According the theory of SWB homeostasis, the single most important thing about SWB is that it is naturally positive (Cummins & Nistico, 2002; Cummins, 2009, in press). This assertion is consistent with the findings from all publications investigating the construct, and all data collected from the Australian Unity Wellbeing Project (Cummins, Eckersley, Pallant, Van Vugt & Misajon, 2003). Further, Cummins and colleagues hypothesise that the positivity of SWB has evolutionary bases, and as such only positive SWB (PWI-A > 50%SM) may be viewed as adaptive. Consequently, all normal SWB set-points must exist in the positive (satisfied) sector of dissatisfied-satisfied continuum range of values (50-

100%SM). In these terms, if depression represents the loss of normal levels of positive adaptive wellbeing, then the presence of depression should evidence homeostatic failure and indicate a maladaptive state. Therefore, it is unlikely that depression will exist in combination with a PWI-A of 70 points or above, whereas it should be ubiquitous when PWI-A is less than 50 points.

Results from initial analyses revealed a number of apparent inconsistencies with these theoretical propositions and predictions. However, it is important to note that such an interpretation relies on the assumption that the DASS21-D is valid in terms of the presumed categories of depression. According to the SWB and DASS21-D (11-point response format) scores:

a. When SWB is greater than 80 points, 10.3% of people report some level of depression (mild, moderate, severe or extremely severe) and 5.05% of the sample combines at least moderate depression with a SWB of greater than 80 points.

b. When SWB lies between 70-80 points, 23.9% of people report some level of depression and 14.2% report moderate or above depression.

c. When SWB is less than 50 points, 12.2% of people report a depression rating of normal.

Although these findings appear to highlight substantial inconsistencies with theoretical predictions from homeostasis theory, the gross prediction of an inverse relationship between SWB and depression was supported. Moreover, in accordance with expectation from homeostasis, results demonstrated that the risk for depression increases substantially at a level of SWB <50%SM. However, the proportion of people reporting depression in the context of SWB scores of greater than 70 points far exceeds expectation. Alternatively, the proportion reporting depression with SWB less than 50 points is far less than what is theoretically expected. These results are not easily explained within homeostasis theory as it is currently articulated. It is possible that a more fine-grain analysis of the relationship between SWB and DASS21 depression may yield further insights into whether this is a flawed prediction from homeostasis theory or whether it has

some other explanation. This analysis was undertaken in the second part of this study.

Hypothesis Four

In confirmation of the operation of homeostatic processes to maintain normal levels of SWB, SWB scores will demonstrate a curvilinear relationship with increasing depression scores that is consistent with Figure 3.

Results partly supported this hypothesis. An examination of PWI-A scores corresponding to depression score groups revealed an inverse relationship. However, the rate at which PWI-A falls was lowest at depression scores of 15.1-18 (moderate) through to 24.1-27.0 (severe), before falling rapidly at higher levels of depression severity.

These findings support the propositions made with SWB homeostasis theory regarding the presence of a curvilinear relationship between SWB and increasing levels of challenge (see Figure 3). However, there were no significant differences between consecutive PWI-A means directly preceding, throughout, or after this moderate-severe 'plateau'. Further, although plotting PWI-A means against the DASS21-D during and after the 'plateau' revealed two apparently different slopes as predicted by theory, these slopes did not differ significantly.

In combination, these two findings fail to confirm the existence of a plateau effect. However, it is possible that failure to verify statistically the presence of a plateau, and therefore the operation of homeostatic processes to resist change, has a methodological explanation. This failure could be due to using DASS21 depression to represent a challenge to SWB, when it may actually function as an indicator of SWB. This proposition will be explained in the context of SWB homeostasis theory.

Levels of SWB as Homeostasis is Challenged

Cummins (2009, in press) has proposed that the critical evidence for the operation of homeostatic processes will be the verification of a curvilinear relationship between levels of SWB and increasing levels of challenge consistent

with Figure 3. Specifically, these researchers proposed that as the strength of challenge to homeostasis progressively increases, mean PWI-A scores evidence a smooth downward trajectory until the lower threshold for homeostasis is reached at about 70 points.

The explanation they offer is that, at any point in time, the exact location of SWB within the set-point range is a probability statement determined by the balance of challenge and homeostatic support. As progressively higher levels of challenge are applied to the homeostatic system, the probability increases that SWB will inhabit the lower reaches of the normative range. As this happens, SWB increasingly becomes bonded to the lower threshold, where it appears to hold steady, thereby demonstrating a plateau effect. However, as the level of challenge continues to strengthen, the homeostatic defence system is eventually overwhelmed and fails, evidenced by a marked drop in SWB.

These researchers also cite evidence for this homeostatic model from an earlier investigation into SWB and depression conducted by Davern (2003, unpublished). These cited results show that SWB demonstrated a curvilinear relationship with depression across increasing depression scores, and homeostatic failure was found to correspond to a depression rating of moderate using the DASS21. In this context, depression was seen to represent the challenging agent, with low depression scores reflecting a low level of homeostatic challenge and high depression scores reflecting a very strong challenge. However, it may be possible that demonstration of a curvilinear relationship between SWB and depression will depend on the nature of the constructs measured within depression scales, and whether these items represent a measure of *challenge* to homeostasis or a measure of outcome (indicator variables).

Causal *versus* Indicator Variables

Investigators in the field of QoL research have drawn a distinction between *causal* (challenge), and *effect* (Fayers & Hand, 1997) or *indicator* (Fayers & Hand, 2002) variables. These two types of variables differ in the relationship they have with the concept being measured. That is, *causal* variables affect the level of SWB, while the changing level of SWB itself is an *indicator*

variable. In these terms, factors such as disease states, symptoms, and treatment side-effects (while indicators of that particular disease) may *cause* reduction in QoL for those people experiencing them, however, the reverse need not apply (Fayers & Hand, 1997). For example, a high level of a symptom, such as sleep disturbance, may imply that a person's SWB is likely to be poor, but a poor level of SWB does not imply that a person experiences disturbed sleep. Sleep disturbance may therefore be viewed either as a *causal* variable of SWB or as reflecting a condition of challenge.

Importantly, variables such as depression, particularly if measured through parameters that are essentially the inverse of positive affect (e.g., anhedonia), are not like sleep disturbance and are not causal. Depression, measured in this way, is a measure of outcome, and so is an indicator variable, just like SWB. Most particularly, if the measures of depression reflect the psychological consequences of the loss of normal levels of positive mood (e.g., hopelessness), then such measures of outcome may be virtually indistinguishable from SWB.

DASS21 depression is measured through such parameters, and therefore scores may represent indicators of SWB. In these terms, it may be reasonable to expect that SWB and DASS21 depression will not necessarily evidence a curvilinear relationship, but rather, they will demonstrate an inverse linear relationship. This is precisely the finding in the current study. Despite a slowing in decreasing levels of SWB, and the apparent emergence of a plateau effect, there was no statistical evidence of a significant plateau in SWB levels across increasing DASS21 depression score increments.

Consistent with the above proposal, it is possible that the trend towards a plateau in SWB means across increasing levels of depression may have resulted from the effect of including two more cognitive symptoms as items (self-deprecation and devaluation of life) in the predominantly affective (anhedonia, inertia, hopelessness, dysphoria and lack of interest/involvement) seven item DASS21-D scale. Results support this suggestion, indicating that of the seven scale items, these two items (self-deprecation and devaluation of life) had mean scores least like the reciprocal of the PWI-A, and that their trajectory across

increasing depression score increments differed significantly from that of the PWI-A.

This result is in line with the notion that the items measuring self-deprecation and devaluation of life are more cognitive, and therefore are less driven by HPMood. In these terms, these two items function more like causal than indicator variables, and therefore, may be more likely to reflect measures of challenge than the affective items. As such, the inclusion of these two cognitive items in the DASS21-D scale may have contributed to the trend towards a slight plateau in SWB scores seen across increasing depression scores. Notwithstanding this, the finding that depression scores increase by 69.30 percentage points across the depression rating of normal to extreme, whereas the PWI-A scores decrease by 28.15 percentage points, suggests that SWB is resisting change, and may provide different evidence for homeostasis as the proposed controller of SWB.

Further research is required to investigate the proposition that depression, like SWB, is a measure of outcome. Moreover, if depression may be viewed as the outcome of a loss of the normal positive HPMood, then it may be possible that measures of SWB represent indices of depression. This suggestion has potentially important definitional implications and gives rise to a number of additional hypotheses regarding the relationship between depression and SWB, and the utility of SWB as novel measures of depression. These hypotheses will be presented in and form the basis for the research conducted in the next study in this thesis.

CHAPTER 7: STUDY 2 INTRODUCTION, AIMS AND HYPOTHESES

The results from Study 1 provide a range of evidence that, in combination, are consistent with homeostasis as the proposed mechanism of SWB control. Specifically, results indicated that SWB is highly stable, and therefore predictable. Further, that SWB resists change in the context of variables that reflect a challenge to homeostasis, evidenced by a visual plateau in SWB levels despite increasing levels of challenge. Additionally, results demonstrated that SWB levels fell markedly directly after the 'plateau effect', supporting the notion that homeostasis has failed, and is no longer in control of SWB. Moreover, the results from Study 1 strengthen the argument that, in line with a dimensional understanding of affective experience, depression may represent the outcome of the loss the normal positive sense of HPMood subsequent to the failure of homeostasis.

These results raised further questions. Specifically, if depression represents the outcome of the loss of HPMood, do SWB measures represent indices of depression? Further, do SWB measures have utility as screening tools for depression? Clearly, the investigation of these questions relies on the reliability and validity of SWB and depression measures. Thus, the major aim of this study is to investigate the comparative validity of three established, traditional self-report measures of depression, and three measures of SWB, as putative novel measures of depression.

In keeping with a dimensional framework, the depression measures used in this study comprise three traditional self-report depression scales. These are the BDI-II (Beck et al., 1996), DASS21-D (Lovibond and Lovibond, 1995), and the HADS-D (Zigmond & Snaith, 1983). The novel measures of depression are HPMood (Cummins et al., 2009, in press), the PWI-A (International Wellbeing Group, 2006), and an item reflecting GLS (International Wellbeing Group, 2006). For description of these instruments, see Measures. Additionally, this study aims to examine a number of theoretical propositions made within SWB homeostasis theory.

To address these research aims, a number of hypotheses are proposed relating to the relationship between SWB and depression, and the ability of SWB and depression measures to identify cases of depression in a sample of general population Australian adults. The first four hypotheses pertain to the relationship between SWB and depression, and the reliability and validity of scales used in this study. The final hypothesis relates to the utility of SWB measures as depression indices.

Specifically, it is hypothesised that:

1. In line with previous research and results from Study 1, there will be a moderate to strong inverse relationship between the measures of SWB and depression.

2. Confirming the findings from Study 1 using a 4-point DASS21-D scale, a rating of DASS21-D mild depression will be associated with the failure of homeostasis at the population level (SWB Mean <70 points).

3. The mean scores for HPMood, PWI-A and GLS will approximate each other.

4. In line with scale content differences between the depression measures, the mean scores for the DASS21-D and the HADS-D will not differ significantly from each other, although these two scales will differ significantly from the BDI-II. The rationale for these differences has been provided in the review of these instruments in Chapter 2.

5. All measures will perform well as screening tools for depression and in detecting depression in the context of research, however, depression scales will perform better than SWB measures.

The statistical analyses required to test these hypotheses are explained in detail in the method chapter to follow (see Statistical Analyses).

CHAPTER 8: STUDY 2 METHOD

The Australian Unity Wellbeing Index (AUWI)

The AUWI is an ongoing research project arising from a partnership between Australian Unity and the Australian Centre on Quality of Life at Deakin University. Through quarterly, and presently biannual, cross-sectional telephone surveys and subsequent longitudinal surveys completed in written form, the AUWI comprises a subjective index of Australians' satisfaction with their lives and their life in Australia. The project commenced in April 2001 and to date, 21 cross-sectional and seventeen longitudinal surveys have been conducted.

Ethics Approval

The current study was approved by Deakin University Human Research Ethics Committee in March 2008 for a period of three years ending March 2011 (see Appendix 2).

The data used in this study are drawn from longitudinal survey 15 (L15) of the AUWI. L15 was designed by me specifically for this study to investigate the comparative validity of three traditional self-report measures of depression and three measures of SWB (see Measures). Appendix 3 contains a copy of this written questionnaire, together with a copy of the plain language statement.

Participants

The current total sample (N=559) comprises a longitudinal cohort from an originally demographically representative sample of Australian adults from the AUWI. Of this sample, 58.2% are female with an age range for men and women combined of eighteen to 88 years (M55.79; SD=14.77). The majority of the sample (65.2%) is married, 9.2% had never married, 6.4% are in a *de facto* relationship, 7.7% are widowed, and the remaining participants are either separated (3.6%), or divorced (7.7%). With respect to work status, 40.2% of the sample is in full-time paid employment, 10.6% is involved in full-time home or family care, 0.4% is engaged in full-time volunteer work, 2.3% is studying full-time and 1.8% is unemployed. The remainder of the sample are retired (30.4%) or

semi-retired (8.9%). Many participants have additional commitments such as part-time paid employment (19.9%), part-time volunteer work (22.0%), and part-time study (5.7%). Household income ranges from < $15K per annum (5.0%) to >$500K (0.4%) per annum, with the greatest percentage of participants (28.4%) reporting an annual household income of $31K-$60K.

Measures

As all measures used in the current study have been presented and discussed in detail previously, only a brief overview of each scale is provided here. The rationale for selecting each of these scales has been described previously. The three depression scales use their original 4-point response scale in order to avoid the potential for methodological issues arising from converting cut-scores to %SM values. This appears important given that each will be used as the depression criterion in separate subsequent ROC analyses.

Beck Depression Inventory – Second Edition

The BDI-II (Beck et al., 1996) measures depression as a psychobiological construct. This scale is included in the current study as it comprises the most widely used measure of self-report depression, and is generally accepted as the 'gold standard' self-report depression inventory.

The BDI-II consists of 21 items presented in a multiple-choice format. Each item statement reflects a specific cognitive, affective, or physiological symptom of depression consistent with DSM-IV diagnostic criteria for an episode of Major Depression. Respondents are instructed to circle the number next to the statement that best describes the way they have been feeling over the past two-week period, including that day. Item statements are rated zero to three, where zero reflects an absence of the depressive symptom and three reflects high intensity of the symptom. BDI-II total scores are calculated by summing the ratings for the 21 items and range from zero to 63. Higher total scores reflect greater severity of symptoms.

Threshold values, or cut-scores recommended by Beck et al. (1996) are outlined in Table 28 below, and are used in analyses as grouping variables for BDI-II depression severity.

Table 28
BDI-II Threshold Scores for Categories of Depression Severity in Clinical Settings (*N*=101)

Total Score	Depression Severity	M	*SD*
0-13	No or minimal depression	7.65	5.9
14-19	Mild	19.14	5.7
20-28	Moderate	27.44	10.0
29-63	Severe	32.96	12.0

The BDI-II has been found to have high internal consistency, with alpha coefficients of .92 and .93 in outpatient and college groups respectively, and test-retest reliability over a one-week period of .93 (Beck et al., 1996).

Depression, Anxiety and Stress Scales

The DASS (Lovibond & Lovibond, 1995) is a 42-item self-report measure of depression, anxiety and stress designed to allow maximum discrimination between the scales, particularly between the constructs of depression and anxiety.

In this study, the depression scale of the DASS short-form, the DASS21-D, was selected for inclusion in the survey, as it is a measure of the psychological symptoms unique to depression. As the DASS21-D does not measure the physiological symptoms of depression that might manifest consequent to the core psychological symptoms of depression, it may represent a purer measure of depression. Additionally, as the manner in which depression is conceptualised within the DASS21-D (psychological), differs from that of the BDI-II formulation of depression (psychobiological), inclusion of this scale allows for the comparative validity of both scales to be examined.

The DASS21-D scale contains seven items constructed to reflect the depressive symptoms of anhedonia, inertia, hopelessness, dysphoria, lack of interest/involvement, self-deprecation, and devaluation of life. Respondents are instructed to circle the number (zero to three) that best indicates how much each statement applied to them over the past week. The depression scale total score is calculated by summing the ratings for the seven items and multiplying this result by two, resulting in a minimum scale score of zero and a maximum scale score of 42 (Lovibond & Lovibond, 1995a). As the DASS21 is a half-length version of the DASS (42 items), doubling of the summed depression items allows for the direct comparison of scores to full scale depression scores, and enables these to be interpreted by reference to the normative values for the full scale. Higher total scores reflect greater frequency/severity of the negative emotional state.

Table 29 summarises the threshold, or cut-scores for categories of depression severity presented within the DASS Manual (Lovibond & Lovibond, 1995a) and used within this study.

Table 29
Categories of DASS Depression Severity

Depression Severity	Total Score	*Z* score	Percentile
Normal	0-9	<0.5	0-78
Mild	10-13	0.5-1.0	78-87
Moderate	14-20	1.0-2.0	97-95
Severe	21-27	2.0-3.0	95-98
Extremely Severe	28+	>3.0	98-100

Alpha values for the DASS21 scales have been found to be: Depression 0.81; Anxiety 0.73; and Stress 0.81 (Lovibond & Lovibond; 1995). The DASS21-D will be referred to simply as the DASS21-D in subsequent text, tables and figures.

Hospital Anxiety and Stress Scales

The HADS (Zigmond & Snaith, 1983) is a brief self-report questionnaire designed to identify *caseness* (possible and probable cases) of depression and anxiety disorders. In this study, only the HADS-D of the HADS is used. The HADS-D was selected for the current study as it is widely used, and represents an alternate measure of the psychological symptoms of depression and as such, expands the range of depression measures available for comparative analyses.

The HADS-D comprises seven items designed to reflect the anhedonic state. Respondents are asked to select the statement that comes closest to the way they have been feeling over the past week. Examples of item statements are ‘I feel cheerful’ and ‘I have lost interest in my appearance’. Response choices differ between items and are scored from zero to three, where zero reflects an absence or low incidence of the depression state being measured (e.g., Not at all, Very Seldom or Hardly at all) and three reflects a higher incidence of that state, or absence of positive emotional experiences (e.g., Often, Definitely or Most of the time).

HADS-D total sores are calculated by summing the seven items; minimum score = 0, maximum score = 21. Although there is no single, generally accepted cut-off score used in interpreting the HADS (Hermann, 1997), Snaith and Zigmond (1994) suggest that depression scores may be interpreted in the following manner. Scores of 7 or less identify non-cases, 8 to 10 mild cases, 11 to 15 moderate cases, and 16 or greater severe cases. These cut-off scores will be adopted in this study as measures of depression severity.

The HADS has been found to have adequate internal consistency, ranging between .60 and .30 for the depression items. Correlations between the HADS and other well-known depression and anxiety inventories range from .67 to .77refs. Results from initial reliability and validity tests indicated that item and scale scores were not affected by physical illness (Zigmond & Snaith, 1983).

Homeostatically Protected Mood

As proposed by Cummins et al. (2009, in press), HPMood reflects the biologically determined positive mood that is hard-wired for each individual and provides the motivation for behaviour across the lifespan. As found by Tomyn (2008), HPMood is summarised by three positive affect items that correspond to the pleasant-activated (happy or alert) and pleasant-deactivated (content) quadrants of the affective circumplex (Russell, 2003). In this study, participants are asked to indicate how each of the following items describes how they generally feel each day on a unipolar end-defined 11-point scale anchored by zero (Not at all) and ten (Extremely). Items are 'How content do you generally feel?', 'How happy do you generally feel?', and 'How alert do you generally feel?'. Scores for the three items are averaged to yield the HPMood score, which is then converted to a percentage point scale from zero to 100 (%SM).

Global Life Satisfaction

GLS represents the most general, abstract, overall rating of LS. It is most commonly measured through the question, 'how satisfied are you with your life as a whole?' In the current study, the response format is a unipolar 11-point end-defined scale anchored by zero to ten, where 0 = *not at all satisfied*, and 10 = *very satisfied*. Research has demonstrated that this measure is reliable (Larsen, Diener & Emmons, 1985), exhibits moderate stability and appropriate sensitivity to changing life circumstances (Eid & Diener, 2004) and has been found to be particularly consistent in Western countries (Cummins, 1995, 1998, 2003). While GLS is an excellent measure of SWB, it is not as reliable as multi-item measures (Cummins, 2003).

The Personal Wellbeing Index – Adult Version

PWI-A (International Wellbeing Group, 2006) represents a first-level deconstruction of satisfaction with life as a whole. It adopts a domain-level approach in which each item measures satisfaction with a different life area, and where each domain explains unique variance in GLS.

The PWI-A comprises eight items of satisfaction that correspond to different QoL domains. These domains are: standard of living, health, achieving in life, relationships, safety, community-connectedness, future security, and religiosity/spirituality.

In this study, respondents are asked to rate their satisfaction with each domain on a unipolar end-defined 11-point scale anchored by zero (not at all satisfied) and ten (very satisfied) with a neutral mid-point of five. Scores for the seven domains are averaged to yield the PWI-A score, which is then converted to a percentage point scale from zero to 100 (%SM).

The PWI-A has sound psychometric properties. A maximum variation of 3.1 percentage points in SWB has been found between the 21 surveys of the Australian population to date (Cummins, 2009, in press). Cronbach's alpha for Australian samples lies between .71 and .85, and test-retest reliability over a one- to two-week period has demonstrated an intra-class correlation coefficient of .84 (Cummins & Lau, 2005).

Procedure

As the procedure for this study is the same as Study 1, only a brief overview will be provided here. A more detailed description of the procedure can be found in the Method section of Study 1.

After obtaining approval from Deakin University Research Ethics Committee for this study, longitudinal survey 15 within the AUWI project, described previously, was designed by me for the purpose of these investigations. A copy of L15 is available in Appendix 3.

A call centre conducts telephone interviews for the initial cross-sectional surveys from which the longitudinal sample for this study were recruited. The longitudinal participants within the AUWI project (described previously) due for annual follow-up via a written questionnaire were identified according to unique numerical identification codes. These codes were sent to Australian Unity, where they were matched to participants' names and address for the purpose of mailing participants the questionnaire. In this way, participants' confidentiality is

maintained as questionnaires returned to Deakin University contain no information allowing for direct identificantion of participants, rather questionnaires have a unique identification code printed onto each page of the questionnaire. Participants may withdraw from the study at anytime, by either not returning the questionnaire, or by noting this on the returned questionnaire.

Statistical Analyses

The aim of the following statistical analyses is to investigate the comparative validity of three widely used self-report depression inventories, and three putative and novel measures of depression. These traditional measures are the BDI-II (Beck et al., 1996), the DASS21-D (Lovibond & Lovibond, 1995a) and the HADS-D (Zigmond & Snaith, 1983) and the novel measures comprise measures of SWB, namely HPMood (Cummins, 2009 in press), the PWI-A (International Wellbeing Group, 2006) and a single GLS item (International Wellbeing Group, 2006).

In the absence of formal diagnosis of Major Depression or Dysthymia through a clinical or structured interview, the traditional measures of self-report depression will each comprise the criterion for depression in separate ROC analyses. As the way in which depression is conceptualised, defined, and operationalised by the BDI-II, DASS21 and HADS-D differ substantially, using these inventories as the criterion for depression affords the opportunity to investigate the differential performance of both traditional and novel indicators of depression across disparate conceptualisations of depression.

The statistical analyses attempt to address this research aim in the following manner. First, the reliability of the three depression scales (BDI-II, HADS-D and DASS21-D) and three SWB measures (HPMood, PWI-A and GLS item) are measured as internal consistencies (Cronbach's alpha). Convergent validity is calculated as Pearson's correlations. Next, means and standard deviations are presented and compared with those reported in the technical manuals for each scale.

However, as the BDI-II, the HADS-D and DASS21-D were originally constructed using 4-point response scales with differing scale anchors, and the

SWB measures use a zero to ten response scale, the key to determining comparable depression and SWB scale scores in terms of their means and standard deviations was to convert all data to a standard form. This conversion makes it look as though all scales had been rated on a zero to 100 point scale. Stokes and Cummins (2006) have outlined the process of converting raw scores into the standard zero to100 scale format in the PWI-A Fourth Edition Manual (International Wellbeing Group, 2006).Values derived from this process are called %SM. Scores on the six measures were calculated as %SM scores according to the formula in the PWI-A Manual as follows:

$$\frac{X - k^{min}}{k^{max} - k^{min}\ x\ 100}$$

X = The score or mean to be converted

k^{min} = The minimum score possible on the scale
i.e. If a scale is score from 1 to 5, then $k^{min} = 1$
If a scale is score from -5 to +5, then $k^{min} = -5$

k^{max} = The maximum score possible on the scale
i.e. If a scale is score from 1 to 5, then $k^{max} = 5$
If a scale is score from -5 to +5, then $k^{max} = +5$

Lastly, for criterion validity, the sensitivity, specificity for each measure of depression and SWB is investigated. These measures are compared using ROC analyses (Lusted, 1971; Metz, 1978) to examine which scale demonstrates superior performance in identifying depression overall, for the purpose of research, and for clinical screening purposes. The criteria used to ascertain scale superiority for these contextual categories is explained subsequently in Results, ROC Analyses. Additionally, ROC analyses will assess the relative validity of each measure of depression, when depression is conceived as a categorical syndrome comprising somatic, affective and cognitive signs and symptoms (BDI-II), as essentially an affective and cognitive dimensional disturbance of mood (DASS21-D), and by a scale initially designed to measure depression in the context of physical illness (HADS-D).

CHAPTER 9: STUDY 2 RESULTS

Data Screening and Examination of Assumptions

Prior to analysis, data were screened for missing values, and accuracy of data entry. Missing values were not replaced in the data set; rather, these were dealt with by retaining all cases and excluding cases pairwise from analyses as recommended by Pallant (2007). Several univariate outliers were detected, however, comparison of mean scores on these variables with corresponding means trimmed at the upper and lower 5% revealed that none of these outliers significantly influenced mean scores on variables germane to this study (BDI-II, DASS21-D, HADS-D, HPMood, PWI-A and GLS). Consequently, univariate outliers were retained and analyses were conducted on the original data. Data were also assessed to ensure no violation of normality, linearity, homoscedasticity multicollinearity and singularity. For normality, SPSS descriptives identified negative skews in SWB scales and positive skews in depression scales. As the skewness for all variables were within the acceptable range of -7.0 to 7.0 (Cohen & Cohen, 1983), no transformations were undertaken. Multicollinearity and singularity were assessed for all variables. As expected, some correlations between the depression measures and between measures of SWB were high ($r > 0.70$). However, as the research aim in the current study is to investigate the comparative validity of three measures of depressions and three measures of SWB, all variables were retained for analyses.

Comparisons between independent ROC Curves were performed using MedCalc Version 10 statistical Software Package. All other statistical analyses in this study were performed using SPSS 17.0 Graduate Package software.

Reliability and Convergent Validity

Internal Consistency and Inter-correlations

For internal consistencies, correlations between the scales, means, and standard deviations reported below, the sample comprised 559 participants as described previously. These participants represented a response rate of 31.94%. from the sample who received the longitudinal survey.

The internal consistency of all five scales is excellent: Cronbach's alpha for the BDI-II was .89; the HADS-D, .77; the DASS21-D, .89; HPMood, .86; and the PWI-A, .83. As the measure of GLS comprised a single item, no internal consistencies are reported.

The relationships between the six measures are investigated using Pearson product-moment correlation coefficients. These correlations are summarised in Table 30.

Table 30

Pearson Product-Moment Correlations between Established (BDI-II, HADS-D and DASS21-D) and Novel Measures (HPMood, PWI-A and GLS) of Depression (N=559)

		1	2	3	4	5
1	BDI-II	-				
2	HADS-D	.78**	-			
3	DASS21-D	.72**	.65**	-		
4	HPMood	-.62**	-.59**	-.63**	-	
5	PWI-A	-.59**	-.55**	-.56**	.78**	-
6	GLS	-.57**	-.56**	-.61**	.81**	.76**

**Correlation is significant at the 0.01 level (two-tailed)

Inspection of Table 30 reveals that all correlations between the six scales are large (r=.50 to 1.0) (Cohen, 1988, pp. 79-81) ranging from -.55 to .78. All correlations between measures are significant (p=.000) at the 0.01 level. As expected, the largest correlations are found within each conceptual group.

The largest correlations between SWB and depression measures occur between the DASS and HPMood (-.63), the BDI-II and HPMood (-.62) and the DASS and GLS (.-61). The smallest correlations, still considered large, occur between the HADS and the PWI-A (-.55), the HADS and GLS (-.56) and the

DASS and PWI-A (-.56). The traditional and novel measures were all inversely related as expected.

Means and Standard Deviations

As the total scores for SWB are calculated as %SM, only depression means and standard deviations are presented in their original form in Table 31. Eta squared values reported subsequent to testing the difference between means are interpreted using the guidelines proposed by Cohen (1988, pp. 284-287) as follows: 0.01 = small effect, 0.06 = moderate effect, 0.14 = large effect.

Table 31
Means and Standard Deviations for BDI-II, DASS21-D, HADS-D, PWI-A, HPMood and GLS Total Scale Scores (N=559)

Scale	*M*	*SD*	*M* (%SM)	*SD* (%SM)
BDI-II	7.55	6.97	11.98	11.06
DASS21-D	6.07	7.29	14.36	17.36
HADS-D	3.09	2.88	14.71	13.71
HPMood	-	-	75.21	14.61
PWI-A	-	-	74.66	13.80
GLS	-	-	75.62	16.15

Depression Scale Means and Standard Deviations (Original Format)

Inspection of Table 31 reveals that for the depression measures in this sample:

1. The BDI-II was normed using psychiatric outpatient samples and one college student sample. As expected, the mean and standard deviation for this general population sample is substantially lower than the BDI-II psychiatric outpatient sample mean (-14.90) and student sample means (-5.01).

2. The DASS21-D mean is very similar to that reported for the normative sample in the DASS Manual (Lovibond & Lovibond, 1995). Using the original scores, the current mean is slightly lower (-0.48) and the standard deviation slightly higher (+0.32) although this is not significant $t(558) = -1.56$, $p = 0.118$.
3. The HADS-D mean and standard deviation are both lower, -0.59 and -0.19 respectively than the sample mean and standard deviation reported by Crawford and colleagues (2001). The difference between means is significant $t(558) = -4.845$, $p = 0.000$, although the magnitude of this difference (mean difference = -0.59, CI: -0.83 to -0.35) is small (eta squared = 0.4). As there is no official normative data reported for the HADS-D, the mean score on the HADS-D from a non-clinical demographically representative sample of more than 1700 British adults used as the HADS normative mean. This, in conjunction with the finding that the difference between means was of a small magnitude, indicates that the HADS-D and the 'normative' sample mean and standard deviation reported by Crawford are generally comparable.
4. Means reported in this sample for the BDI-II, DASS21-D, and HADS-D correspond with depression severity levels of minimal (BDI Score=0-13), normal (DASS21-D Score=0-9), and non-case (HADS-D Score=0-7) respectively, and these are in line with those reported in the technical manuals for each scale.

Taken together, these findings indicate that the means and standard deviations for depression scales in the current study are comparable to the normative data means and standard deviations for each scale. As such, total scale scores in the current study may be interpreted using the recommended cut-scores for categories of depression severity reported by the authors of the scale.

The means and standard deviations for all three SWB measures lie within the respective Australian adult normative ranges (PWI-A 73.43%SM to 76.43%SM; GLS 75.20%SM to 79.10%SM).for SWB generated from the 21 surveys from the AUWI project (N=42,464).

Depression Scale Means and Standard Deviations (%SM)

From Table31, the greatest difference between the standardised depression scale means occurs between the HADS-D and BDI-II (2.73 points), and the smallest difference occurs between the DASS21-D and HADS-D (0.35 points). Consistent with these differences, and as expected, the results of three t-tests indicate that there is no significant difference between the mean scores for the DASS21-D (M=14.36, SD=17.36) and the HADS-D (M=14.71, SD=13.71), $t(558) = 0.463$, $p = 0.644$ (two-tailed). The magnitude of the differences in means (mean difference = 0.26, 95% CI: -0.86 to 1.38) is very small (eta squared = 0.0003). However, there is a statistically significant difference between the mean scores on the DASS21-D and the BDI-II (M=11.98, SD=11.06), $t(558) = 4.82$, $p = 0.0005$ (two-tailed) and between the HADS-D and BDI-II, $t(558) = 7.55$, $p = 0.0005$ (two-tailed) although the magnitude of the differences in the means (mean difference = 2.46, CI: 1.46 to 3.46) is small-moderate (eta squared = 0.04) for the DASS21-D and BDI-II and moderate (mean difference = 2.72, CI: 2.01 to 3.43, eta squared = 0.09) for the HADS-D and BDI-II. This is in line with differences between the content of the scales. That is, the BDI-II measures depression through a combination of cognitive, affective and physiological symptoms, the DASS21-D and HADS-D measure only the psychological symptoms of depression.

For SWB Measures, the %SM means and the standard deviations appear to approximate each other, with the mean score for the three measures varying less than one percentage point (0.96 points). In line with these small observed mean differences, the results of three t-tests to compare the mean scores on the three measures of SWB are not significant at the 0.01 level. This attests to the reliability and validity of these measures.

Interestingly, examination of the mean scores within categories of depression severity, reveals that the mean score for each of the SWB measures falls from above 70 points at a depression rating of Normal/No depression to lie below the proposed lower threshold for homeostasis (<70 points) at a depression rating of mild for the BDI-II and DASS21-D. As the HADS-D has only one cut-score for caseness, this occurs at a level corresponding to a possible/probable case. These are: for BDI-II mild depression, HPMood M=62.47 SD=11.20, PWI-

A *M*=62.28 *SD*=13.17, GLS *M*=63.40 *SD*=13.94; for DASS21-D mild, HPMood *M*=67.12 *SD*=12.55, PWI-A=64.38 *SD*=12.68, GLS *M*=68.18 *SD*=15.14; and for HADS-D Possible/Probable Case, HPMood *M*=55.30 *SD*=15.85, PWI-A *M*=55.73 *SD*=16.15, GLS *M*=54.62 *SD*=17.60. These results, using depression scales with 4-point response formats, are in line with the finding that levels of SWB fell below the lower homeostatic threshold at a level of mild when depression was measured using the 4-point DASS21-D scale.

Comparative Validity of Scales according to Differing Criteria for Depression

To investigate the comparative validity of three depression scales and three measures of SWB, three separate ROC analyses (Metz, 1978) are performed.

ROC Analysis

ROC analysis is a technique used for visualising, organising, and selecting classifiers based on their performance (Fawcett, 2005). Originally used in signal detection theory (Egan, 1975), ROC graphs have utility in analysing the behaviour of diagnostic systems (Swets, Dawes & Monahan, 2000) and are widely used in medicine to evaluate the accuracy of a test to discriminate diseased cases from normal cases (Swets, 1996; Swets, Dawes & Monahan, 2000). In this study, the accuracy with which scales are able to discriminate depressed cases from normal cases will be evaluated.

ROC Curve and Area under the Curve (AUC).

In essence, ROC analyses produce two-dimensional graphs in which the true positive rate (sensitivity) is plotted against the false positive rate (specificity) for successive test scores. Each point on the ROC plot represents a different sensitivity/specificity pair corresponding to a particular decision threshold (cut-score). In combination, these points form what is known as the ROC Curve. A test with perfect discrimination (100% sensitivity, 100% specificity) possesses an ROC Curve that passes through the upper left corner of the ROC plot (Zweig & Campbell, 1993). In this context, the closer the ROC Curve is to the upper left corner of the plot, the higher the overall accuracy of the test, and the greater the

area under curve (AUC). Therefore, the value for the AUC represents a convenient indicator of overall test performance.

Cut-scores.

Cut-scores are specific test scores that distinguish between cases and non-cases of the variables under investigation. In this study, cut-scores refer to specific scale total scores on each of the five measures whose test performance is under examination. Each cut-score has an associated sensitivity and specificity level. As the sensitivity and specificity levels of the scale will vary across successive scale total scores, identification of the optimum cut-score for a test may depend on the purpose for using the test, the priorities of the clinician or researcher, and the unique characteristics of the sample (Beck et al., 1996).

Specifically, for screening purposes, sensitivity is prioritised over specificity as higher sensitivity means fewer false negatives (undetected cases of depression). This may be a desirable first step in a two-stage clinical screening process, when the purpose is to detect the maximum number of persons with depression. However, high sensitivity also increases the rate of false positives (cases erroneously classified as depressed) and may result in the unnecessary referral of otherwise healthy individuals for further assessment. To limit the occurrence of false positives, while maintaining adequate sensitivity, a cut-score that demonstrates maximal sensitivity *and* a specificity of at least 60% will be used to derive the first cut-score reported for each scale in this study.

For research studies in which specificity is the paramount criterion, Beck et al. (1996) recommend raising the cut-scores to reduce the number of false positives, although to what level they should be raised is not specified. This second cut-score prioritises specificity over sensitivity, and is defined as maximal specificity + sensitivity $\geq$ 60%. This cut-score is recommended for research purposes and comprises the second cut-score reported in the results below.

The third cut-score reported for each scale, the maximal Youden Index (Youden, 1950), is one of the most frequently used indices of overall test performance. Arguing that it was not a statistical matter to decide the relative seriousness of the two types of diagnostic error, Youden proposed the index as a

measure to reflect equal valuation of sensitivity and specificity. As such, the maximal Youden Index does not weight one type of error over another, rather it aims to characterise the optimal performance of a test, as calculated by summing sensitivity and specificity-1.

The reporting of these results will adhere to the following procedures. In the event that the parameters used for identifying cut-scores in contextual categories (Screening, Research and maximal Youden Index) results in either the sensitivity or specificity level of a scale falling below 50%, no cut-score will be reported for the scale in the category in which this occurred. Additionally, if obtaining a sensitivity level of 60% results in the specificity level falling below the sensitivity level for research purposes, no cut-score will be reported for *Research* and the scale will not be recommended for use in the context of research studies. Similarly, if meeting the 60% minimum for specificity for screening purposes results in the sensitivity falling below the specificity level, no cut-score will be reported for *Screening* and the scale is not recommended for use as a screening tool. Finally, if adherence to the category inclusion parameters results in the same specificity and sensitivity pair being identified as the optimal levels for a scale in more than one contextual category, then the same cut-score will be reported for these categories.

ROC Samples

The sample used in each ROC analysis comprises a different sub-set of the total sample described previously. Subsamples comprise those participants who reported a level of depression less than or equal to that of mild depression according to the cut-scores of the scale used as the depression criterion in each analysis. For example, in the first ROC analysis, the BDI-II was used as the criterion for depression, where a total BDI-II score classified as *No* depression represented a negative case, and a BDI-II score classified as mild depression represented a positive case. Cases with higher levels of depression were excluded from the analysis.

ROC Analysis 1: Comparative Validity Using BDI-II Mild Depression as Depression Criterion (*N*=523)

This ROC analysis was performed on a sample selected as a BDI-II total score of zero to nineteen representing the range of values from no depression to mild depression. Corresponding scale total scores for the DASS21-D, HADS-D, HPMood, PWI-A and GLS were entered into the analysis as test variables. According to the depression criterion, there were 50 (9.56%) positive cases (BDI-II total score=14-19) and 473 (90.44%) negative cases (BDI-II total score=0-13).

As depression and SWB are inversely related, scores on measures of HPMood, the PWI-A and GLS were re-coded so that high scores on the three SWB measures reflect low levels of HPMood, less satisfaction with life domains, and less satisfaction with life as a whole. Re-coding these variables inverted the ROC Curves for SWB measures, thereby facilitating the comparison between the ROC Curves for the SWB measures and the measures of depression.

The ROC Curves of the five measures for BDI-II mild depression are depicted in Figure 13.

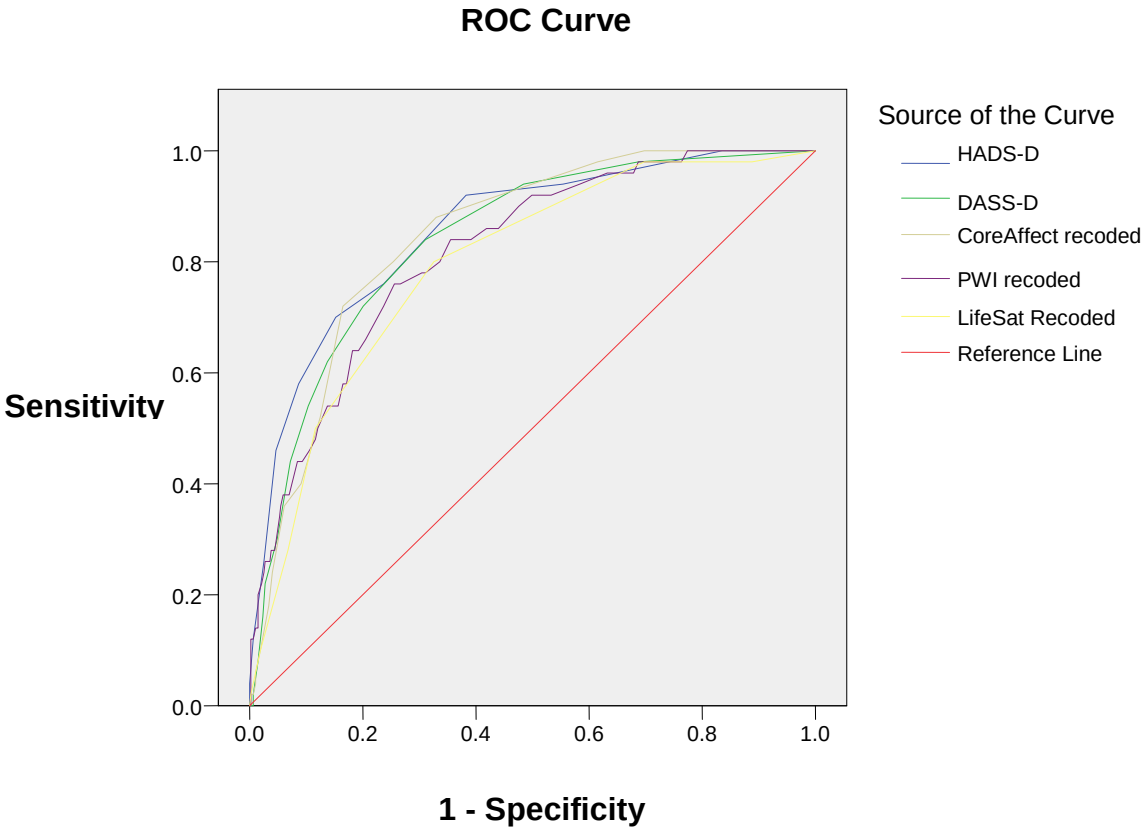


Figure 13. Mild depression: ROC Curves for cut-off points (total scale scores) of HADS-D, DASS21-D, HPMood, PWI-A and GLS (*N*=523). BDI-II mild depression was the criterion for calculating sensitivity and specificity of the scales

From Figure 13, it remains unclear which of the five measures possesses the largest AUC as ROC Curves intersect repeatedly. However, it appears that the ability of these measures to discriminate positive from negative cases of BDI-II depression may fall loosely into two groups. One group of measures, comprising the DASS21-D, HADS-D and HPMood, perform more similarly to each other and better than the other group comprising the PWI-A and GLS item. Areas under the curves are 0.856 (*SE*=0.027; 95%*CI*=0.802-0.909) for HADS-D, 0.843(*SE*=0.027; 95%*CI*=0.796-0.891) for HPMood, 0.839 (*SE*=0.024; 95%*CI*=0.785-0.892) for DASS21-D, 0.816 (*SE*=0.029; 95%*CI*=0.759-0.874) for PWI-A, and 0.794 (*SE*=0.031; 95%*CI*=0.733-0.855) for GLS. These AUC values demonstrate that each measure was significantly better than chance (AUC= 0.5) at discriminating

between respondents with and without mild depression (p=0.000). Of the five measures, the HADS-D achieved the largest AUC (0.856), appearing to marginally outperform all other scales (HPMood+0.013; DASS21-D +0.017; PWI-A +0.04; GLS +0.062), although the differences between AUCs are small.

In order to test whether these differences were statistically significant, ten pairwise comparisons between AUCs for independent ROC Curves were performed using MedCalc Version 10.3 statistical software programme. As these tests involved multiple comparisons, the Holm-Bonferroni method (Holm, 1979) was used to calculate the alpha level for each test. This method was selected as it controls the family-wise error rate for all k hypotheses at level α in the strong sense and each intersection is then tested using the simple Bonferroni test (Holm, 1979). In this way, the Holm-Bonferroni method is a uniformly more powerful correction than the Bonferroni correction (Abdi, 2007).

For ten pairwise comparisons, the alpha level ranged from 0.005 to 0.05. Results of these comparisons and the alpha level at which each comparison was assessed (α H-B) are presented in Table 32.

Table 32

Pairwise Comparison of Areas under Independent ROC Curves

	HPMood					DASS21-D					PWI-A					LS				
	Dif.	*SE*	*z*	*p*	α H-B	Dif.	*SE*	*Z*	*p*	α H-B	Dif.	*SE*	*z*	*p*	α H-B	Dif.	*SE*	*z*	*p*	α H-B
HADS-D	0.013	0.04	0.34	0.73	0.025	0.017	0.04	0.45	0.66	0.017	0.040	0.04	1.01	0.31	0.007	0.062	0.04	1.51	0.13	0.005
HPMood						0.004	0.04	0.11	0.91	0.05	0.027	0.04	0.68	0.50	0.008	0.049	0.04	1.19	0.23	0.006
DASS21-D											0.023	0.04	0.61	0.54	0.01	0.045	0.04	1.15	0.25	0.006
PWI-A																0.022	0.04	0.52	0.60	0.0125

Table 32 summarises the results for each of the ten pairwise comparisons including the difference between AUCs (Dif.) the standard error (SE), the *z* score, *p* value, and the Holm-Bonferroni alpha level (α H-B). Inspection of these results reveals that:

1. There were no significant differences between the AUCs for the HADS-D, DASS21-D, HPMood, the PWI-A and GLS. This result suggests that the overall ability of instruments to detect depression did not differ.
2. The standard error for each comparison remains constant (*SE*=0.04) across all comparisons suggesting that the results from these comparisons are likely to generalise across samples.

Nevertheless, examination of the operating characteristics of each scale at specific points (cut-scores) determined by the parameters described previously for Screening, Research and following calculation of the Youden Index, reveals differences between the scales' sensitivity and specificity levels.

The ability of each scale to detect cases from non-cases of depression is summarised in Table 33 as three cut-scores, each with its corresponding sensitivity and specificity level. In Table 33, the first cut-score presented for each scale (Screening) is recommended for initial screening purposes and is derived by identifying the cut-off point that provides maximal sensitivity in the context of a specificity level of at least 60%. The second cut-score for each scale (Research) is recommended for research purposes and is derived by identifying the cut-off point that provides maximal specificity in the context of sensitivity levels of at least 60%. The third cut-score (maximal Youden Index) represents the optimal classification performance of the scale. This cut-score is derived by identifying the maximal result from summing sensitivity and specificity-1 for each possible cut-point on the scale.

Table 33

BDI-II Depression: Operating Characteristics of the HADS-D, HPMood, DASS21-D, PWI-A and GLS (*N*=523)

Scale	Screening (maximal sensitivity & specificity ≥60%)			Research (maximal specificity & sensitivity ≥ 60%)			Youden Index maximal (sensitivity + specificity – 1)		
	Cut-score	Sensitivity	Specificity	Cut-score	Sensitivity	Specificity	Cut-score	Sensitivity	Specificity
HADS	≥3	92.0	61.7	≥5	70.0	84.8	≥5	70.0	84.8
HPMood	≥30	88.0	67.0	≥32	72.0	83.5	≥25	88.0	67.0
DASS21-D	≥5	84.0	68.9	≥9	62.0	86.3	≥5	84.0	68.9
PWI-A	≥24	84.0	60.9	≥32	64.0	81.8	≥32	64.0	81.8
LS	≥30	80.0	67.4	-	-	-	≥30	80.0	67.4

Inspection of Table 33 reveals that across the three categories of cut-scores, the sensitivity of the five scales varies 30.0%, ranging from 62.0% for the DASS21-D (Research) to a high of 92.0% for the HADS (Screening). Specificity levels were slightly lower, varying 24.6% from a low of 61.7% for the HADS (Screening) to 86.3% for the DASS (Research).

For *Screening* purposes, all measures demonstrate an adequate level of utility. Of the five measures, the two traditional depression measures displayed the highest and third highest sensitivity levels (HADS-D, 92.0%; DASS21-D, 84.0%). This result is not unexpected, given that these measures were designed specifically to measure the core constructs of depression as conceptualised by the authors of the scale. What is surprising, however, is that for a small compromise in sensitivity (-4%), HPMood, a measure of SWB, demonstrates high utility as a screening tool for BDI-II, with the second highest sensitivity level (88.0%) in the context of a specificity level that is 5.3% higher than that of the HADS-D. Given that limiting the number of false positives, is extremely important for effective resource management in a clinical setting, HPMood may offer greater overall utility than the HADS-D as a first step in a two stage screening process.

For *Research* purposes, the GLS item did not meet the category inclusion criteria and for this reason no cut-score is reported and this item is not recommended for use in the context of research. Of the remaining 4 measures, the DASS21-D displayed the highest specificity (86.3%), although this was associated with a sensitivity level of <65% (62.0%). Whereas, the HADS-D and HPMood demonstrated high specificity levels of 84.8% and 83.5% respectively while also maintaining sensitivity levels of greater than or equal to 70% (HADS-D, 70.0%; HPMood, 72.0%). In these terms, the decision as to which scale possesses the greatest utility in a research context is not transparent. Rather, these cut-scores and their associated sensitivity and specificity levels may inform the researcher and facilitate selection of the instrument to best meet their research priorities and environment. Notably, in both research and screening contexts, HPMood, a measure of SWB, appears to perform as well as the HADS-D, and better than the DASS21-D, in detecting cases of depression within specific contextual parameters.

The optimal test performance of the five measures is summarised by the maximal Youden Index. Inspection of the sensitivity and specificity levels within this cut-score category reveals that the HADS-D and PWI-A both display higher specificity (HADS-D, 84.8%; PWI-A, 81.8%) than sensitivity (HADS-D, 70.0; PWI-A, 64.0%). As such, these two scales may be classified as most effective for research purposes. The other three scales (HPMood, DASS21-D and LS) possess higher sensitivity than specificity levels, for HPMood sensitivity = 88.0%, specificity 67.0%; for DASS21-D sensitivity = 84.0%, specificity =64.9%; LS sensitivity = 80.0%, specificity = 67.4%, and therefore, their optimal performance demonstrates these three measures may have greater utility in screening for cases of depression.

Taken together, these results indicate SWB measures, particularly HPMood, may have as much utility in research contexts and perhaps greater utility in clinical settings in differentiating cases from non-cases of BDI-II mild depression than two of the most widely used traditional self-report depression inventories.

ROC Analysis 2: Comparative Validity Using DASS21-D Mild Depression as Depression Criterion (*N*=475)

This ROC analysis was performed on a sample selected as a DASS21-D total score of zero to thirteen, representing the range of values from no depression to mild depression. Corresponding scale total scores for the BDI-II, HADS-D, HPMood, PWI-A and GLS were entered into the analysis as test variables. According to the depression criterion, there were 44 (9.26 %) positive cases (DASS21-D total score=10-13) and 431 (90.74%) negative cases (DASS21-D total score=0-9).

As in the previous ROC analysis, scores on measures of HPMood, the PWI-A and GLS were re-coded so that high scores on these three measures reflected low levels of HPMood, less satisfaction with life domains, and less satisfaction with life as a whole. This inverted the ROC Curves for the measures of SWB, facilitating the comparison between the ROC Curves for SWB and depression.

The ROC Curves of the five measures for DASS21-D mild-moderate depression are depicted in Figure 14.

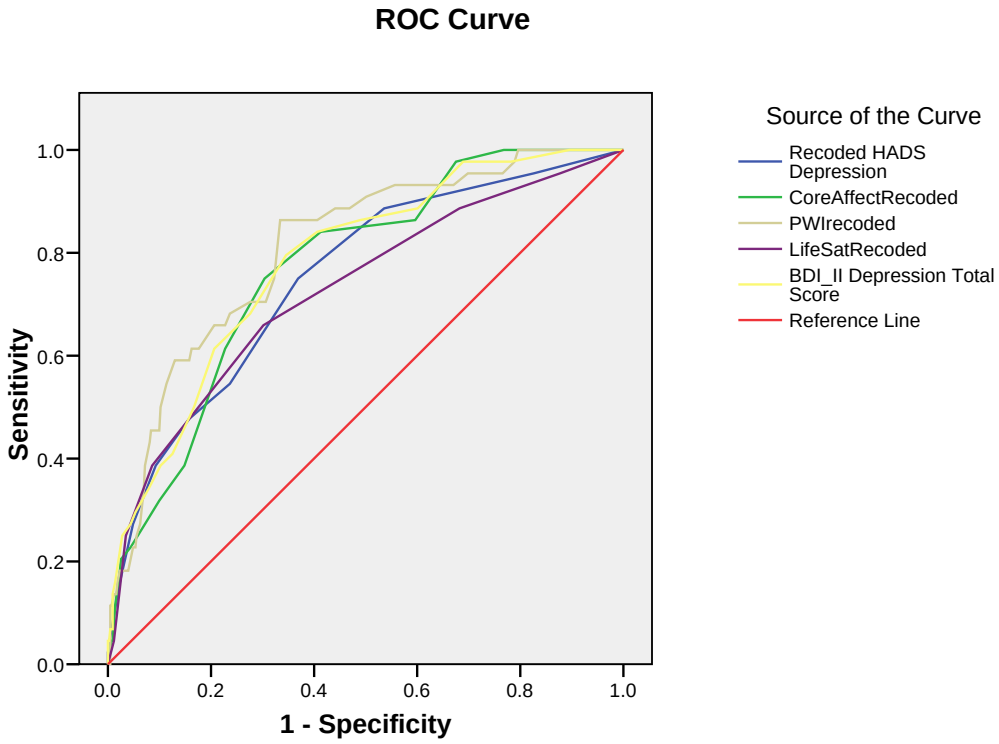


Figure 14. Mild depression: Receiver operating characteristic curves for cut-off points (total scale scores) of HADS-D, BDI-II, HPMood, PWI-A and GLS (*N*=523). DASS21-D mild depression was the criterion for calculating sensitivity and specificity of the scales

From Figure 14, it remains unclear which of the five measures possesses the largest AUC as ROC Curves repeatedly intersect across the plot. Areas under the curves are 0.807 (*SE=0*.033; 95%*CI*=0.743-0.871) for PWI-A, 0.782 (*SE*=0.034; 95%*CI*=0.715-0.849) for the BDI-II, 0.770 (*SE*=0.033; 95%*CI*=0.704-0.835) for HPMood, 0.750 (*SE*=0.033; 95%*CI*=0.674-0.826) for HADS-D, and 0.726 (*SE*=0.043; 95%*CI*=0.642-0.811) for GLS. These AUC values demonstrate that each measure was significantly better than chance (AUC = .5) at discriminating between respondents with and without mild depression (p=0.000). Of the five measures, the PWI-A achieved the largest AUC (0.807) and the smallest AUC was obtained by the GLS item (0.726). It is notable that the difference between the largest AUC and smallest AUC is less than 0.1 (0.081). As

such, it is unlikely that such small differences between the AUCs would be statistically significant. Nevertheless, the differences observed between the AUCs were tested via ten pairwise comparisons of the AUCs for independent ROC Curves using MedCalc Version 10.3 statistical software programme. As these tests involved multiple comparisons, the Holm-Bonferroni method (Holm, 1979) was used to calculate the alpha level for each test. For ten pairwise comparisons, the alpha level ranged from 0.005 to 0.05. Results of these comparisons and the alpha level at which each comparison was assessed are presented in Table 34.

Inspection of the results in Table 34 reveals that there were no significant differences between the AUCs for the HADS-DBDI-II, HPMood, PWI-A and GLS. This result indicates that ability of SWB measures to discriminate cases with depressive symptoms from cases with no depressive symptoms did not differ significantly from two established, traditional self-report measures of depression.

As there are no significant differences between the overall performances of the five scales being tested, examination of the sensitivity and specificity levels may assist researchers and clinicians alike in the selection of the most appropriate self-report inventory to identify cases of depression for their specific purposes. Three cut-scores and their associated sensitivity and specificity levels are presented for each scale in Table 35. The first cut-score is recommended for screening purposes, the second for research purposes and the third, the maximal Youden Index, represents the optimal performance of test. The derivation of these cut-scores have been described previously (see ROC Analysis 1).

Table 34

Pairwise Comparisons of Areas under Independent ROC Curves. Depression Criterion = DASS21-D Mild Depression (*N*=475)

	BDI-II					HPMood					HADS-D					LS				
	Dif.	*se*	*z*	*p*	α H-B	Dif.	*se*	*z*	*p*	α H-B	Dif.	*se*	*z*	*p*	α H-B	Dif.	*se*	*z*	*p*	α H-B
PWI-A	0.025	0.05	0.52	0.60	0.0125	0.037	0.05	0.79	0.43	0.008	0.057	0.05	1.12	0.26	0.006	0.081	0.05	1.49	0.14	0.005
BDI-II						0.012	0.05	0.25	0.80	0.05	0.032	0.05	0.62	0.54	0.01	0.056	0.06	1.02	0.31	0.006
HPMood											0.02	0.05	0.37	0.71	0.025	0.044	0.05	0.81	0.42	0.007
HADS-D																0.024	0.06	0.41	0.68	0.017

Dif.=Difference between AUCs; α H-B=Alpha level for each comparison derived through Holm-Bonferroni method

Table 35

DASS21-D Mild Depression: Operating Characteristics of the HADS-D, HPMood, DASS21-D, PWI-A and GLS (*N*=523)

	Screening (maximal sensitivity & specificity ≥60%)			Research (maximal specificity and sensitivity ≥ 60%)			Youden Index maximal (sensitivity + specificity – 1)		
Scale	Cut-score	Sensitivity	Specificity	Cut-score	Sensitivity	Specificity	Cut-score	Sensitivity	Specificity
PWI-A	≥23	86.4	61.7	≥31	61.4	83.8	≥23	86.4	61.7
BDI-II	≥7	79.5	65.4	≥9	61.4	79.4	≥7	79.5	65.4
HPMOOD	≥25	75.0	69.6	≥28	61.4	77.3	≥25	75.0	69.6
HADS-D	≥3	75.0	63.1	-	-	-	≥4	54.5	76.3
Life Sat.	-	-	-	≥30	65.9	69.8	≥30	65.9	69.8

From Table 35, across the three categories of cut-scores, the sensitivity of the five scales varies 25%, from 61.4% for the PWI-A, BDI-II, and HPMood (Research) to 86.4% for the PWI-A (Screening and maximal Youden Index). Specificity levels show a similar range, varying 22.1%, from a low of 61.7% for the PWI-A (Screening) to a high of 83.8% for the PWI-A (Research).

For *Screening* purposes, the GLS item did not meet the category inclusion criteria and for this reason no cut-score is reported and this item is not recommended for use a screening tool for DASS mild depression. The remaining four measures (PWI-A, BDI-II, HPMood and HADS-D) demonstrate an adequate level of utility with sensitivity levels of >70% and an associated specificity level of at least 60%. Of these four measures, the PWI-A displays the highest sensitivity (86.4%) levels; however, this occurs in the context of the lowest specificity level (61.7%). While the sensitivity levels for the BDI-II and HPMood were somewhat lower than for the PWI-A, -6.9%, and -11.9 % respectively, the corresponding specificity levels for these two measures are higher +3.7%, and +7.9% respectively. As such, the BDI-II and HPMood may represent viable options for use in clinical settings when limiting the rate of false positives is important.

For *Research* purposes, no cut-score is reported for the HADS-D, and this scale is not recommended for use in the context research, as it did not meet the category inclusion criteria. The PWI-A demonstrates the highest specificity level of 83.8, followed by the BDI-II (79.4%), and HPMood (77.3%). As the three highest specificity levels correspond to the same level of sensitivity (61.4%), it appears that the test performance of the PWI-A for research purposes is superior to that of the other four measures in this context. While the GLS item demonstrates the lowest specificity (69.8%), this reflects the item's optimal level of test effectiveness (maximal Youden Index, sensitivity=65.9, specificity=69.8%).

The maximal Youden Index summarises the optimal test performance of the five measures. Three of the measures, the PWI-A, BDI-I and HPMood display higher sensitivity than specificity levels (PWI-A sensitivity=86.4%, specificity=61.4%; BDI-II sensitivity=79.5%, specificity=65.4%; HPMood

sensitivity=75.0%, specificity=69.6%), and may be classified as most effective as screening tools. Whereas, the HADS-D (sensitivity=54.5, specificity=76.3%) and the GLS item (sensitivity=65.9%, specificity=69.8%) possess higher specificity than sensitivity and, therefore, their optimal performance indicates that these scales may have greater utility in identifying cases of depression in research settings than clinical settings.

From these results, it appears that the PWI-A, a SWB measure of respondents' satisfaction with specific life domains, may have greater utility in both clinical and research contexts in detecting cases of DASS21-D mild depression than the other measures of SWB and two of the most widely used self-report depression inventories.

ROC Analysis 3: Comparative Validity Using HADS-D Mild Depression as Depression Criterion (*N*=547)

This ROC analysis was performed on a sample selected as a HADS-D total score of zero to ten representing the range of values from no case of depression to possible/probable case of depression. As the HADS-D scale originally contained only one cut-score, eight, to delineate non-cases from possible/probable cases of depression, the sample was limited to a HADS-S score of ten to exclude cases with more severe depression symptoms in keeping with the sample composition of the previous two ROC analyses. Corresponding scale total scores for the BDI-II, DASS21-D, HPMood, the PWI-A and GLS were entered into the analysis as test variables. According to the depression criterion, there were 39 (7.13%) positive cases (HADS-D total score = 8-10) and 508 (92.87) negative cases (HADS-D total score = 0-7).

The ROC Curves of the five measures for HADS-D mild depression are depicted in Figure 15.

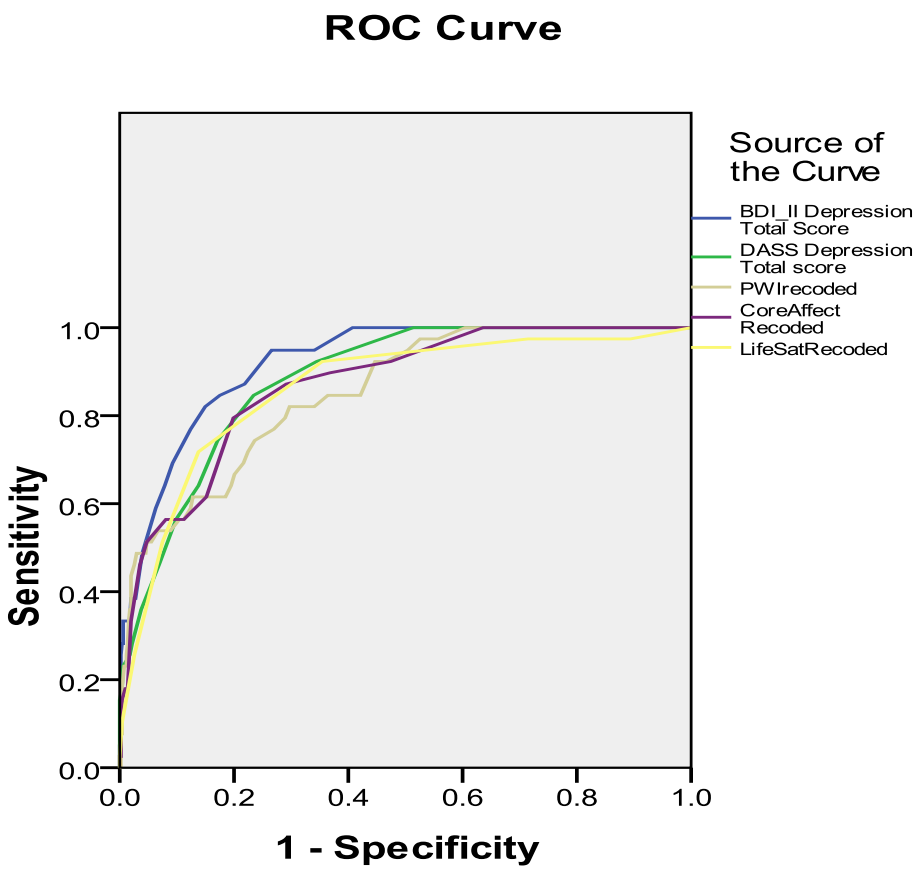


Figure 15. HADS-D possible/probable case: ROC curves for cut-off points (total scale scores) of BDI-II, DASS21-D, HPMood, PWI-A and GLS (*N*=547). DASS21-D mild depression was the criterion for calculating sensitivity and specificity of the scales

While the ROC Curves within Figure 15 intersect repeatedly, it appears that the BDI-II may possess the largest AUC of the five measures. Areas under the curves are 0.919 (*SE*=0.017; 95%*CI*=0.885-0.953) for BDI-II, 0.882(*SE*=0.022; 95%*CI*=0.839-0.925) for DASS21-D, 0869 (*SE*=0.027; 95%*CI*=0.816-0.922) for HPMood, 0.861 (*SE*=0.031; 95%*CI*=0.800-0.922) for GLS, and 0.852 (*SE*=0.029; 95%*CI*=0.795-0.909) for the PWI-A. These AUC values are high, and higher than those found in the previous two analyses are. Each measure performed significantly better than chance (AUC = .5) at discriminating between possible/probable cases and non-cases of depression according to the HADS-D (p=0.000). The difference between the largest AUC

(0.919, BDI-II) and smallest AUC (PWI-A, 0.852) is 0.067, and this is in line with the largest and smallest AUC values (<0.1) found in ROC Analyses 1 and 2.

In order to test whether these differences were statistically significant, ten pairwise comparisons of the AUCs for independent ROC Curves are performed using MedCalc Version 10.3 statistical software programme. As these tests involved multiple comparisons, the Holm-Bonferroni method (Holm, 1979) was used to calculate the alpha level for each test. For ten pairwise comparisons, the alpha level ranged from 0.005 to 0.05. Results of these comparisons and the alpha level at which each comparison was assessed are presented in Table 36.

Table 36

Comparison of Areas under Independent ROC Curves using HADS-D Possible/Probable Case as the Depression Criterion

	DASS					HPMood					LS					PWI-A				
	Dif.	*se*	*z*	*p*	H-B α	Dif.	*se*	*z*	*p*	H-B α	Dif.	*se*	*z*	*p*	H-B α	Dif.	*se*	*z*	*p*	H-B α
BDI	0.037	0.03	1.33	0.18	0.007	0.05	0.03	1.57	0.12	0.006	0.05	0.03	1.57	0.12	0.006	0.067	0.03	1.99	0.05	0.005
DASS						0.013	0.04	0.37	0.71	0.017	0.021	0.04	0.56	0.57	0.01	0.03	0.04	0.82	0.41	0.008
CA											0.008	0.04	0.20	0.85	0.05	0.017	0.04	0.43	0.67	0.013
LS																0.009	0.04	0.21	0.83	0.025

From Table 36 it is clear that there were no significant differences between the AUCs for the BDI-II, DASS21-D, HPMood, the PWI-A and GLS. This result indicates that ability of SWB measures to determine when HADS-D depression was present did not differ significantly from two of the most widely used self-report depression scales designed specifically to assess the psychological symptoms of depression (DASS21-D), and to assess the symptoms that comprise the diagnostic criteria for DSM-IV-TR Major Depression (BDI-II).

As there are no significant differences between the overall performance the five classifiers, the sensitivity and specificity levels require examination as a means to aid instrument selection as a function of context and purpose.

Table 37 summarises the recommended cut-scores and their associated sensitivity and specificity for screening and research purposes, and to indicate optimal test performance for each of the five measures.

Table 37

HADS-D Depression: Operating Characteristics of the BDI-II, DASS21-D, HPMood, PWI-A and GLS (*N*=547)

Scale	Screening (maximal sensitivity with specificity ≥60%)			Research (maximal specificity with sensitivity ≥ 60%)			Youden Index maximal (sensitivity + specificity – 1)		
	Cut-score	Sensitivity	Specificity	Cut-score	Sensitivity	Specificity	Cut-score	Sensitivity	Specificity
BDI-II	≥8	95.0	65.9	≥15	64.1	92.1	≥11	84.6	82.5
DASS21-D	≥5	92.3	65.6	≥11	64.1	86.2	≥7	84.6	76.6
HPMood	≥25	90.0	63.2	≥35	61.5	85.0	≥32	79.5	80.1
Life Sat.	-	-	-	≥35	71.8	86.2	≥35	71.8	86.2
PWI-A	≥25	84.6	61.4	≥37	61.5	87.2	≥28	82.1	70.3

Inspection of Table 37 reveals that for *Screening* purposes, the GLS item did not meet the category inclusion criteria and for this reason, no cut-score is reported and this item is not recommended as a screening tool. The remaining four measures all display good utility, and three of these demonstrate a test sensitivity of greater than or equal to 90% (BDI-II, 95.0%; DASS21-D 92.3%; HPMood 90.0). While the PWI-A has the lowest sensitivity, (84.6%), this level is still considered high. In this context in which sensitivity is the paramount criterion, the BDI-II appears to outperform the other measures displaying the highest sensitivity (95.0%) and the highest level of specificity (65.9%) within this cut-score category.

For *Research*, the BDI-II demonstrates the highest level of specificity (specificity=92.1%, sensitivity=64.1%). The difference between the specificity of the remaining four measures is small (2.2%), with the PWI-A achieving the second highest specificity overall (87.2%) and HPMood the lowest (85.0%). While all measures have good utility in this contextual category, it appears that the BDI-II demonstrates the greatest test effectiveness for research purposes.

According to the maximal Youden Index, the optimal performance of three measures (BDI-II, DASS21-D and PWI-A) evidence higher sensitivity than specificity levels, indicating that these scales may be classified as more effective for screening than for research purposes. Of these, the BDI-II has the highest sensitivity/specificity pair (84.6%/82.5%), followed by the DASS21-D (84.6%/76.6%), and the PWI-A (82.1%/70.3%). When the optimal performance of the GLS item and HPMood is summarised, however, these measures demonstrate higher specificity than sensitivity (GLS, 86.2%/71.8%; HPMood 80.1%/79.5%), and indicate that they may be most effective in the context of research.

Taken together, these results suggest that both traditional self-report measures of depression and SWB perform well as screening tools (sensitivities for all measures ≥85%) and in the context of research (specificity for all measures ≥85%) for HADS-D mild depression. Although of the five measures, the BDI-II appears to have the greatest utility for delineating the presence from the absence of HADS-D mild depression in a cohort of general population Australian adults.

Results Summary

The aim of this study was to investigate the comparative validity of three widely used, traditional self-report depression inventories, (BDI-II, DASS21-D and HADS-D) and three putative and novel measures of depression (HPMood, PWI-A and GLS item). Statistical analyses were performed to address this research aim and the results of these allow for the examination of the reliability (Cronbach's alpha), convergent validity (Pearson's product-moment correlations), and criterion validity (ROC Analyses) of each of the six scales. The results from the statistical analyses are summarised briefly below.

The reliability of four measures, the BDI-II, DASS21-D, HPMood and the PWI-A, are excellent (Cronbach's alpha 0.83-0.89). The HADS-D demonstrates good internal consistency (Cronbach's alpha=0.77). No internal consistency is reported for GLS as this measure comprises a single item.

For convergent validity, in line with previous findings, strong (r = >.5) negative correlations are found between the measures of depression and SWB, with high levels of depression associated with low levels of SWB. In addition, there are large positive correlations between the scales within each of the two groups of measures respectively (depression scales, r = .65-.78, and SWB measures, r =.76-.81). All correlations are significant (p = 0.000). In combination, these correlations indicate a strong association among all six scales. While all correlations are high, the highest correlations are among the SWB instruments, suggesting that these instruments are more likely to be measuring the same construct, and have greater construct validity than the depression scales.

The means and standard deviations (original response format) of depression scales (BDI-II, DASS21-D and HADS-D) for the total sample approximate those reported for general population samples in the technical manuals for each scale. The general comparability between the means and standard deviations of depression scales found in this study, together with the comparability of those reported in the technical manuals for each scale, supports the use of the respective cut-scores for categories of depression severity recommended by the authors of the scale. For SWB measures, the means and

standard deviations approximate each other and lie within the respective Australian adult normative ranges for the PWI-A and GLS indicating that results are comparable across studies.

For standardised (%SM) means and standard deviations, the largest mean difference occurs between the BDI-II and the HADS-II (2.73 points), while the smallest difference occurs between the DASS21-D and HADS-D (0.35 points). This is in line with differences in scale content, as the BDI-II is a measure of both the psychological and physiological symptoms of depression, whereas the DASS21-D and HADS-D measure depression as a purely psychological construct. The difference between the standardised means of the SWB measures is smaller than that for the depression measures, varying less than 1 percentage point (0.96 points). This attests to the reliability and validity of these measures.

In the first ROC analysis using BDI-II mild depression as depression criterion (*N*=523), AUC values (AUCs=0.794-0.856) indicate that all measures (DASS21-D, HADS-D, HPMood, PWI-A and GLS) perform at a level significantly better than chance (AUC=0.5) for discriminating between respondents with and without mild depression. While the HADS-D (AUC=0.856, *SE*=0.027, 95%*CI*=0.802-0.909) and HPMood (AUC=0.843, *SE*=0.027, 95%*CI*=0.796-0.891) achieved the greatest AUC values, the difference between the largest and smallest AUC was 6.2%. Ten pairwise comparisons between AUCs for independent ROC Curves reveals that the differences between AUCs are not statistically significant ($p<$*Holm-Bonferroni* α). These findings indicate that the ability of SWB measures to determine when depression was present did not differ significantly from two of the most widely used self-report depression scales designed specifically to assess the symptoms characteristic of, and unique to depression.

Examination of the sensitivity and specificity levels for all measures across categories of cut-scores (Screening, Research and maximal Youden Index), indicate that SWB measures, particularly HPMood, may have as much utility in research contexts and perhaps greater utility in clinical settings in differentiating cases from non-cases of BDI-II mild depression than two of the most widely used traditional self-report depression inventories.

In the second ROC analysis using DASS21-D mild depression as depression criterion (*N*=475), all measures (BDI-II, HADS-D, HPMood, PWI-A and GLS) perform at a level significantly better than chance (AUC=0.5) for discriminating between respondents with and without mild depression (AUCs=0.726-0.807). According to AUC values, the PWI-A (AUC=0.807, *SE=0*.033, 95%*CI*=0.743-0.871), BDI-II (AUC=0.782, *SE*=0.034, 95%*CI*=0.715-0.849), and HPMood (AUC=0.770, *SE*=0.033, 95%*CI*=0.704-0.835) demonstrate the best overall test performance. Although the difference between the best and worst test performance is 8.1%, ten pairwise comparisons between AUCs for independent ROC Curves confirms that the overall ability of tests to discriminate people with depressive symptoms from people with no depressive symptoms does not significantly differ ($p < \alpha$ H-B).

Inspection of the sensitivity and specificity levels for all measures across categories of cut-scores (Screening, Research and maximal Youden Index) indicate that the PWI-A, a SWB measure of respondents' satisfaction with specific life domains, may have greater utility in both clinical and research contexts in detecting cases of DASS21-D mild depression than two other measures of SWB and two of the most widely used self-report depression inventories.

In the third and final ROC analysis using HADS-D mild depression as depression criterion (N =574), With the exception of the PWI-A, AUC values for all measures (BDI-II, DASS21-D, HPMood and GLS) are higher than those found in the previous two analyses (AUCs=0.861-0.919). Each measure performs significantly better than chance (AUC = .5) at discriminating between possible/probable cases and non-cases of HADS-D Depression ($p = 0.000$). In line with the largest and smallest AUC values found in ROC Analyses 1 and 2, AUC values in this analysis vary less than 0.1 (0.067). Ten pairwise comparisons between AUCs for independent ROC Curves indicate there are no significant differences between the overall test performance of the five measures ($p < \alpha$ H-B).

Evaluation of the sensitivity and specificity levels of all measures across cut-score categories (Screening, Research and maximal Youden Index) reveal that both traditional self-report measures of depression and measures of SWB perform

well as screening tools (sensitivities for all measures ≥85%) and in research settings (specificity for all measures ≥85%) for HADS-D mild depression. Although of the five measures, the BDI-II appears to have the greatest utility for delineating the presence from the absence of HADS-D mild depression.

CHAPTER 10: STUDY 2 DISCUSSION

The major aim of this study was to investigate the comparative validity of three established, traditional self-report measures of depression, and three measures of SWB, as measures of depression. The traditional self-report depression scales were the BDI-II (Beck et al., 1996), the DASS21-D (Lovibond & Lovibond, 1995a), and the HADS-D (Zigmond & Snaith, 1983). The novel measures of depression were HPMood (Cummins, 2009, in press), the PWI-A (International Wellbeing Group, 2006), and an item reflecting GLS (International Wellbeing Group, 2006). Additionally, this study sought to test a number of theoretical propositions made within Cummins' theory of SWB homeostasis.

To address these research aims, a number of hypotheses were proposed relating to the relationship between SWB and depression, and the ability of SWB and depression measures to identify cases of depression in a sample of Australian general population adults. The first four hypotheses pertained to the relationship between SWB and depression and the final hypothesis related to the utility of SWB measures as depression indices.

These hypotheses were that:

1. In line with previous research and results from Study 1, there will be a moderate to strong inverse relationship between the measures of SWB and Depression.
2. Confirming the findings from Study 1, a rating of DASS mild depression will be associated with the failure of homeostasis at the population level (SWB Mean <70 points).
3. The mean scores for HPMood, PWI-A and GLS will approximate each other.
4. In line with scale content differences between the depression measures, the mean scores for the DASS21-D and the HADS-D will not differ significantly from each other, although these two scales will differ significantly from the BDI-II. The rationale for these differences has been provided in the introduction.

5. All measures will perform well as screening tools for depression and in detecting depression in the context of research, however, depression scales will perform better than SWB measures.

Results from statistical analyses will be discussed in relation to each of these hypotheses.

Hypothesis One

In line with previous research and results from Study 1, there will be a moderate to strong inverse relationship between the measures of SWB and depression.

Results supported the first hypothesis. There was a strong inverse linear relationship between the three self-report measures of depression and the three measures of SWB, ranging from -.55 between the PWI-A and HADS-D, to -.63 between the DASS21-D and HPMood. This is consistent with the findings from Study 1, and with a substantial body of empirical evidence (Cheung & Bagley, 1998; Davern, 2004; Goldberg & Harrow, 2005; Hansson, 2002; Heady, Kelly & Wearing, 1993; Hong & Giannakopoulos, 1994; Koivumaa-Honkanen et al., 2001, 2004; Lewis et al., 1999; Simpson et al., 1996). In combination, these findings are consistent with the theoretical proposition that depression represents the failure of homeostasis and subsequent loss of the positive sense of HPMood (Cummins, 2009) as detailed in Hypothesis Four. However, determining the level of depression severity at which homeostasis may be seen to fail is not transparent, as is evident from testing the next hypothesis.

Hypothesis Two

In confirmation of findings from Study 1, the mean score for the PWI-A will fall below the lower threshold for group means (<70 percentage points) at a DASS depression rating of mild depression.

The level of depression severity associated with the failure of homeostasis has important implications for the theoretical propositions made within SWB homeostasis theory (Cummins et al., 2009, in press) and for the validity of results

from ROC analyses in the current study. In line with the findings from Study 1 in this thesis, and confirming this hypothesis, PWI-A means fell below the lower threshold for homeostasis for group means (< 70 points) at a DASS depression rating of mild using the 4-point response scale.

In Part A of Study 1, the comparability of depression data collected from surveys containing two versions of the DASS was examined. One version used the original 4-point response scale, the other an end-defined 11-point response scale. It was found that PWI-A means within DASS depression categories were higher (from +2.75 to +8.53 points) using the 11-point response scale than the 4-point scale. Importantly, this difference between the two response formats translates into different diagnostic categories when the SWB scores are projected onto the depression scales.

Using the original 4-point response formats for depression scales, the means for all three measures of SWB (PWI-A, HPMood and GLS) fall below the lower homeostatic threshold (70 points) at a depression rating of mild according to all three of the depression scales (BDI-II, DASS21-D and HADS-D). However, using the 11-point format, the 70-point value of SWB corresponds to a depression rating of moderate.

These findings raise the possibility that the differing levels of depression severity associated with the failure of homeostasis for 4-point and 11 -point scales are an artefact of measurement. That is, these differences are a consequence of the lack of correspondence between the transformed cut-scores for the 4-point and 11 -point response scales used to measure DASS Depression. These differences, their effect on the classification of cases and, therefore, the level of depression associated with the failure of homeostasis, will now be explained and discussed.

As described previously, cut-scores to delineate categories of DASS depression severity were based on the 4-point response format. These cut-scores correspond to five ratings of depression severity, where DASS21-D total scores of zero to nine correspond to a rating of normal, ten to thirteen for mild, fourteen to 20 for moderate, fifteen to nineteen for severe, and 20+ for a rating of extremely severe. Importantly, these cut-scores were verified by examining the agreement

between DASS scores and formal diagnoses of depression through structured clinical interview in clinical and general population samples during the scale's development (Lovibond & Lovibond, 1995a). However, cut-scores have never been established for the 11-point version.

To allow for the comparison of depression scale scores generated by using 4-point and 11-point versions of the DASS, all DASS data must be transformed to a standard format, such as %SM as outlined by Stokes and Cummins in the PWI-A Manual (International Wellbeing Group). This transformation makes it look as though both response scales had been rated on a zero to 100 point scale. Then, in order that the standardised response scores can be interpreted by reference to the cut-scores for categories of depression severity, the original cut-scores must also be transformed to %SM values. This, however, creates the problem.

The assumption of a simple transformation is that the transformed cut-scores for the 4-point will apply equally to the 11-point. Yet this is incorrect. Rather, because the 4-point response scale is a more rudimentary measure than the 11-point response scale (Cummins & Gullone, 2000), there will be a lack of correspondence between the transformed cut-scores and the transformed 11-point categories. Specifically, converting the original cut-scores to the standard form creates discontinuity between the cut-scores for depression ratings. For example, the highest depression score for a rating of DASS normal depression is nine. This corresponds to 21.42%. However, the next highest depression score of 10 (mild), when transformed, yields a standardised value of 23.81%. Thus, the transformation has created a 'dead-space' between the two depression ratings of 2.39 points. Similar between-ratings discrepancies are seen for the other categories of depression.

As the original cut-scores are based on the 4-point response scale, the 'dead-spaces' between transformed depression categories do not affect the classification of 4-point total scores as the scoring formats for total scores and cut-scores are the same. However, the discontinuity between depression ratings does affect the classification of 11-point total scores as some transformed 11-point total scores will fall into the 'dead-space' between the transformed cut-scores for

depression categories due to the expanded response scale. This was precisely the situation encountered in the first study.

In Study 1, all 11-point data (*N*=2329) and 4-point data (*N*=1308) and DASS21-D cut-scores were transformed to %SM values. Following these transformation, it was found that 140 of the 11-point PWI-A scores fell into the 'dead-spaces' between the ratings of depression, with the largest number of these occurring between the depression ratings normal-mild (*N*=74), and mild-moderate (*N*=43). No 4-point PWI-A scores were found between the threshold scores for transformed depression categories.

In order to match the PWI-A scores with the depression categories, the %SM values for depression categories needed to be moved up or down to eliminate these 'dead-spaces'. To enhance scale sensitivity, cut-scores were adjusted downwards to minimise the number of false negatives as recommended by Beck et al. (1996).

The decision to lower the %SM cut-scores had repercussions for the classification of 11-point cases into depression categories, and for the average level of SWB found within depression categories. The classification of cases in a rating of normal (0-21.42 points) was unchanged. The lowest score for a depression rating of mild (23.81%SM) was adjusted downwards to 21.43 points so that the 74 cases that fell between ratings of normal and mild (21.43%SM-23.81%SM) were incorporated into the mild depression category. Likewise, the lowest score for a rating of moderate depression (33.33%SM) was decreased to 30.96%SM so that the 43 cases that fell between the highest score for mild depression (30.95 points) and the lowest score for moderate depression (33.33 points) were incorporated into the moderate depression category. Similar adjustments were made for the remaining cases (*N*=23) across levels of severe-extremely severe depression. Effectively, these adjustments result in lower 11-point total scores within ratings of DASS21-D mild-extremely severe depression and this affects the PWI-A Mean reported across depression categories.

As detailed in Hypothesis One, a strong inverse linear relationship has been found between the SWB and depression, therefore the lower the depression

score, the higher the corresponding PWI-A score. In this context, the lowering of transformed cut-scores means that DASS21-D depression ratings will incorporate lower depression scores than they would in the absence of adjustment. Further, the addition of a substantial number of lower scores to each category of depression will result in lower DASS21-D means, and therefore higher PWI-A means within transformed depression ratings.

Consistent with this proposition, in Study 1, PWI-A means were higher for 11-point relative to 4-point ratings of depression; for mild +3.98 points, for moderate +4.28 points, for severe +8.53 points, and for extremely severe depression +7.12 points. Higher PWI-A means within depression ratings indicates that the average PWI-A score will move below 70 points at more severe level of depression when using the 11-point response scale than when using the 4-point scale. In line with this observation, the results from Study 1 demonstrated that the level of depression associated with failure of homeostasis at a group level (<70 points) corresponded to a rating of mild for 4-point data and moderate for 11-point data. Replication of this finding in the current study using three 4-point depression scales provides strong support for the contention that the differing levels of depression severity associated with the failure of homeostasis for 4-point and 11-point scales are an artefact of measurement.

In these terms, it is possible that the differing levels of DASS depression severity associated with homeostatic failure for 4-point (mild) and 11-point scales (moderate) both have merit. However, as depression scales in the current study use 4-point response scales, these findings support the use of a depression rating of mild to define a positive case of depression in ROC analyses. The next hypothesis concerns the reliability and validity of the SWB measures.

Hypothesis Three

The mean scores for HPMood, PWI-A and GLS will approximate each other.

Results support the hypothesis. According to SWB homeostasis theory, the mean scores for all three measures of SWB should approximate each other, because they are essentially measuring the same construct. Specifically, SWB can

be predominantly characterised as a stable positive mood that is held within a narrow set-point range for each individual by homeostatic processes. As such, SWB may be considered an approximation of HPMood, which is a pervasive, abstract, and positive feeling about the self (Cummins, 2009, in press).

GLS, as measured through the question 'How satisfied are you with your life as a whole?' (Andrews & Withey, 1976), represents the most general abstract overall rating of LS. This question lacks definition with respect to the precise target of the response, making it unlikely to be answered using a systematic cognitive approach (Schwarz & Strack, 1991, 1999). Instead, this question is probably answered via a heuristic to Current Mood (Mackay, 2008), which under normal operating conditions approximates the set-point level of HPMood (Cummins, 2009).

Finally, the PWI-A (International Wellbeing Group, 2006) comprising eight items of satisfaction with different life domains, represents a first-level deconstruction of GLS, and should therefore yield a similar mean score to the GLS mean. Consistent with homeostasis theory, a series of paired samples t-tests confirmed that there were no significant differences between the mean scores of the three SWB measures. Therefore, these findings supported the first hypothesis.

Notably, consistent with the Australian adult normative statistics from the AUWI, the mean score for the PWI-A (74.66%SM) was within the Australian adult normative range of 73.43%SM to 76.43%SM. Additionally, the mean score for GLS (75.62) was also within its normative range of 75.20%SM to 79.10%SM. It is also noteworthy that the mean score for HPMood (75.21) is consistent with the Australian normative ranges for both the PWI-A and GLS. These results confirm the first hypothesis and attest to the validity and reliability of these SWB measures. Additionally, these findings indicate that results from the current study may be compared with those from Study 1, and may generalise across studies using Australian adult general population samples.

Perhaps the most important feature of these results is that they are consistent with the core characteristics of SWB as proposed within the theory of SWB homeostasis. First, with only 5.72 % of the current sample reporting SWB

levels in the negative (< 50 points) sector of the response scale, results support the notion that SWB is naturally positive. Second, as SWB means lie within the extraordinarily narrow Australian adult normative ranges, these results provide further evidence of the remarkable stability of SWB and support the proposition that each individual has a genetically determined set-point that is maintained by homeostatic processes. Finally, as the mean score for HPMood lies within the normative ranges for the PWI-A and for GLS, these results are in line with previous research and the theoretical proposition that HPMood is the dominant affective constituent of SWB (Davern & Cummins, 2004; Tomyn, 2008), and comprises the steady-state, set-point that homeostasis seeks to defend (Cummins et al., 2009). However, self-report depression scales do not demonstrate this level of between measures consistency, as is apparent by testing the next hypothesis.

Hypothesis Four

The mean scores for the depression scales will differ in line with scale content differences. Specifically, the DASS21-D and HADS-D mean will not differ significantly, although means for both scales will differ significantly from the BDI-II mean.

Results generally supported this hypothesis. There were no significant differences between the means of the DASS21-D and the HADS-D. However, the means of both scales were significantly higher (+2.38; +2.73 points) than the BDI-II mean. It seems likely that these differences are a consequence of the specific items chosen to represent depression by each of the scales' authors, and the reporting period for depressive symptoms. These will now be examined in turn.

Differences in Scale Content

The DASS (Lovibond & Lovibond, 1995a) was designed to measure and discriminate between the negative emotional states of depression, anxiety and stress. In so doing, the authors of the scale aimed to further the process of understanding, defining and measuring these constructs. Using a complex empirical research strategy (described previously), only those symptoms that did not overlap with other psychological conditions, namely anxiety and stress, were included in the final depression scale. These symptoms are negative affect and

negative cognitions about the self and one's life, and a loss of positive affect (anhedonia). Therefore, the DASS21-D comprises a measure of the psychological symptoms of depression.

Similar to the psychological formulation of depression offered by the DASS21-D, the HADS-D also comprises items selected specifically to reflect the anhedonic state, since this is probably the core construct of depression (Zigmond & Snaith, 1983). As the HADS-D was originally designed for use in medical outpatient settings, physiological symptoms, such as dizziness and headaches, were excluded as a means to differentiate mood disorders from the secondary effects of physical illness. In these terms, it appears that the DASS21-D and HADS-D attempt to measure only those symptoms unique to depression, and that these symptoms are of a psychological nature. The finding that there was no significant mean difference between the two scales confirms that they have operationalised to measure depression in a similar manner. However, as HADS-D and DASS21-D means were significantly higher (+2.38; +2.73) than the BDI-II mean, it is likely that the HADS-D and DASS21-D differ from the BDI-II in terms of scale content.

Examination of the items comprising the BDI-II reveals that depression is conceptualised as a psychobiological syndrome comprising a set of cardinal signs (anhedonia and depressed mood) and a suite of associated symptoms (somatic and cognitive disturbances). BDI-II scale items were expressly designed by Beck and colleagues (1996) to reflect the DSM-IV diagnostic criteria for an episode of Major Depression and therefore adopt a disease approach to measuring depression. As the BDI-II definition and measurement of depression is commensurate with the most widely accepted medical conceptualisation of depression as a disorder, it is not surprising that this scale is generally accepted as the gold standard self-report measure of depression. In these terms, the inclusion of physiological symptoms in the BDI-II scale distinguishes this scale from the DASS21-D and HADS-D, and this inclusion might reasonably be expected to explain the significant difference found between mean scores. However, this is not entirely correct.

The BDI-II contains items that may be related to physical illness as well as depression. Consequently, it may be expected that a greater number of people will endorse a greater number of scale items as potentially more than one condition is being measured. Thus, it may be reasonable to expect that the BDI-II would yield higher total scores than the DASS21-D and HADS-D, and therefore demonstrate a higher mean score. However, in the current study, the BDI-II mean was significantly lower than the DASS21-D and HADS-D means. This raises the possibility that other factors may have contributed to this finding. Specifically, it is possible that the differences in scale content are attenuated by differences in the reporting periods for symptoms between the scales. Further, the differences between reporting periods may also have important definitional and measurement implications for Depression and possibly, for SWB. These will now be discussed.

Differences in Reporting Periods

For the general public and in psychology, the terms ‘depressed’ and ‘depression’ refer to a lowering of mood (dysphoria) that can vary in severity from a fluctuation in normal mood to an extreme feeling of sadness, pessimism and despondency (Faravelli et al., 2005; American Psychological Association, 2007). In contrast to this view of depression as a continuum of depressed mood, in medicine depression is conceptualised within a categorical framework. Here it represents a syndrome that differs from normal feelings of sadness in terms of symptom duration, quantity, qualitative severity, and presence of manifest other symptoms. This medical definition of depression is used for diagnosing depression as a disorder in the DSM-IV-TR (American Psychiatric Association, 2004).

As previously stated, the BDI-II was designed specifically to be consistent with this medical definition of depression, with scale items selected to reflect all nine symptoms listed as DSM-IV-TR diagnostic criteria for Major Depression. Importantly, the reporting period for symptoms in the BDI-II is the past two weeks, and this is in line with minimum required period for the presence of symptoms for a diagnosis of Major Depression. In conjunction with the number of symptoms, the duration of symptoms comprises the key to distinguishing Major Depression (five of nine criteria, must include anhedonia or depressed mood for

past two weeks) from Dysthymic Disorder (three symptoms for four years), and dysphoria known as subthreshold depression (<5 symptoms, < 2weeks) (Ingram & Siegle, 2002).

From this medical perspective, depression scales that measure fewer symptoms than those listed as diagnostic criteria for depressive disorders, and/or include reporting periods for symptoms of less than two weeks, measure the more commonly experienced state of subthreshold depression (Ingram & Siegle, 2002). In these terms, subthreshold depression is more closely aligned with the lay definition of depression as a ubiquitous negative emotional state and, due to the relaxation of inclusion criteria, is experienced with much greater frequency in general populations samples than diagnosable depressive disorders (Judd et al., 1994; Judd et al., 1997).

As the DASS21-D and HADS-D measure only the psychological symptoms of depression in reference to a reporting period of the past week, these two scales may be viewed as measures of subthreshold depression. Alternatively, since the BDI-II measures depression in a manner consistent with a DSM-IV diagnosis for Major Depression (over the past two weeks), it may be viewed as a measure of clinical or Major Depression. In these terms, given that the frequency of subthreshold depression is manifold higher than clinical depression in general population samples (Judd et al., 1997), and self-report scales measure the frequency/severity of depressive symptoms, the DASS21-D and HADS-D as measures of subthreshold depression should demonstrate higher total scale scores, and evidence higher scale mean scores than the BDI-II as a measure of clinical depression.

This was precisely the finding within the current study; the DASS21-D and HADS-D means were significantly higher than the BDI-II mean. Support for the validity of this interpretation of results is found in the significantly higher frequency of cases classified as normal according to the BDI-II (84.6%) relative to the DASS21-D (77.1%). As previously described in the Results section, the method used by the HADS-D for classifying cases into categories of depression excludes this scale from being used comparatively with the BDI-II and DASS21-D in this instance.

In conclusion, the scale content of depression scales, and the reporting period for symptoms, reflects the framework within which depression is conceptualised (BDI-II, categorical; DASS21-D and HADS-D, dimensional) and defined (BDI-II, medical; DASS21-D and HADS-D, lay/psychological). As such, these differences should be considered and used to guide researchers and clinicians in selecting the most appropriate instrument for their specific purposes. Notwithstanding these definitional implications, manipulation of the reporting period for symptoms may also have implications for the measurement of affect in general and for the relationship between SWB and depression within homeostasis theory.

Reporting periods and the measurement of affective states *versus* affective traits.

Lovibond and Lovibond (1995) assert that the DASS was designed to emphasise states, rather than traits, and suggest that the DASS achieves this by asking the respondent to rate the degree to which each symptom was experienced over the past week. These researchers argue that limiting the reporting period for item statements to this period, minimises respondent reports of *traits* (what the respondent characteristically feels, or experiences). While not stated explicitly, it is likely that these authors use traits to mean personality traits.

Interpreting Lovibond and Lovibond's use of the term traits in the context of personality appears reasonable. Previous research has found that people with high levels of neuroticism typically experience higher levels of negative affect and negative cognitions than their extraverted counterparts (Rusting & Larsen, 1997). In this context, it is possible that Lovibond and Lovibond (1995a) were attempting to minimise the probability of incorrectly classifying people, such as those with high levels of neuroticism, as depressed, when their normal level of mood may not represent dysfunction. Within the context of SWB homeostasis theory, this group of people might correspond to those with a low set-point for HPMood.

Interestingly, Lovibond and Lovibond also state that DASS items are not directly applicable to momentary emotional states (what the respondent feels at precisely that point in time) as some items refer to experiences and situations

outside of the testing context. In combination with the previous distinction made between states (over the past week) and traits (what the respondent typically experiences or characteristically feels), it appears that Lovibond and Lovibond have unintentionally made a distinction between three levels of affect (momentary, state, trait) as opposed to two (state, trait), which was their stated purpose. The proposition of three levels of affect is noteworthy and is consistent with the affective framework within the theory of SWB homeostasis. Further, given that depression is a disturbance of mood, SWB homeostasis theory may offer a more relevant, and more parsimonious, approach to defining traits in the context of depression, and for understanding and differentiating affective states from affective traits.

These ideas can be elaborated within the context of SWB homeostasis as follows:

1. There are three levels, or layers, of affect. These are: i) Momentary Emotions (transient, reactive, classified as transient affect); ii) Current Mood (duration hours/days/weeks, classified as state-level affect); and iii) HPMood (trait-like properties, biologically determined, enduring, stable, object-free, classified as trait-level affect).

2. Consistent with Lovibond and Lovibond's suggestions (1995), it is proposed that these three layers of affect may be differentially measured by manipulating the reporting period for scale items. To measure Momentary Emotions, the respondent could be asked to rate the degree to which he/she experienced each item right now, for Current Mood, the last few days, or past week, and for HPMood , how the respondent generally feels every day.

3. SWB measures typically cast items in the *trait* form by referring the respondent to what he/she generally feels about their life (GLS), specific aspects of their life (PWI-A), and the extent to which they generally feel happy, alert, and content each day (HPMood). Conversely, depression scales attempt to minimise measuring *traits,* casting items in the state form by referring the respondent to the way they have been feeling in relation to

particular depressive symptoms over the past week (DASS21-D and HADS-D) or past two weeks including today (BDI-II).

4. Under normal operating conditions, Current Mood should approximate the set-point for HPMood. Therefore, under normal operating conditions, only two levels of affect are experienced or felt by the individual, and for the purpose of measurement, Current Mood (state) and HPMood (trait) will not differ despite the use of different reporting period for affect items.

5. As previously stated, under normal operating conditions state affect (Current Mood) should approximate the set-point for trait affect (HPMood). At a group level, this corresponds to a normative range for SWB of 70-80 points. Inverting this normative range to reflect depression corresponds to depression scores of 20-30 percentage points. It is notable that the cut-scores for a normal rating of depression according to the three depression scales using 4-point response formats generally fall within this range (BDI-II 20.64 points; DASS21-D 21.42 points; HADS-D 33.33 points). This provides support for the congruence between levels of affect under conditions of homeostatic equilibrium.

6. In the event of strong, prolonged negative challenge, homeostasis eventually fails, and the positive sense of HPMood is lost. At this point, an individual loses contact with HPMood, Current Mood reflects the challenging agent, and the individual's dominant affective experience will be negative. It is proposed that the loss of positive HPMood and the consequent separation of Current Mood from HPMood is the essence of depression. In these terms, HPMood and Current Mood (depression) may be differentially measured, although they will continue to evidence a strong inverse linear relationship as found in this thesis and in previous research.

In this context, it is interesting to speculate whether altering the reporting period for SWB measures to reflect a one- to two-week period may increase their sensitivity and specificity as indices of depression and further illuminate the relationship between SWB and depression. Future research in this area is required

to test the viability of this proposition. Notwithstanding this possibility, the congruence between the findings in the current study and SWB homeostasis theory provide strong support for the argument that SWB measures have utility as indices of depression as is evident from testing the next hypothesis.

Hypothesis Five

All measures will perform well in differentiating cases of depression from normal cases, however, depression scales will perform better than SWB measures

Results generally supported the final hypothesis. The accuracy of tests in detecting the presence of depression ranged from moderate (AUC = 0.70 – 0.90) to high (AUC > 0.90) for BDI-II, DASS21-D and HADS-D depression, with all measures performing at a level significantly better than chance. However, the sensitivity and specificity levels of measures varied across contexts (Screening and Research) and according to the depression criterion used. Nevertheless, without question, the most interesting finding within this study was that some measures of SWB performed as well as traditional depression scales in differentiating cases of mild depression from normal cases, and in some instances, performed better than depression scales specifically designed for this purpose.

These findings are consistent with the theoretical proposition that depression represents the loss of the normal, positive sense of HPMood. Further, the current findings have implications for depression screening in primary health care settings. The discussion will now turn to the overall performance of tests, the utility of SWB as screening tools, and the implications of these findings for the propositions made within SWB homeostasis theory.

Do Traditional Depression Scales Represent Superior Indices of Self-report Depression?

Conclusions drawn from any investigation regarding the criterion validity of instruments depend on the validity of the reference criterion. While this study may be criticised for not employing a ‘gold standard’ diagnostic clinical interview as the criterion for depression, it is important to highlight that the main aim of this study was not to investigate the utility of SWB as indices of MDD. Rather, this

study aimed to investigate the comparative validity of three established, widely used, self-report depression scales, and three putative and novel measures of depression. All measures were assessed against differing formulations of self-report depression, as the reference criteria for depression.

This approach is consistent with a dimensional conceptualisation of affective experience. In this conceptualisation, mental disorders, such as depression, are viewed as existing on a continuum with normality (Eyesenck, Wakefield & Friedman, 1983; Widiger, 1997), and Major Depression represents the end-point along a continuum of depressive symptomatology, as opposed to a structurally distinct entity (Ruscio & Ruscio, 2000). Additionally, a dimensional framework is consistent with the understanding of affect within the theory of SWB homeostasis, and the proposition that depression represents the loss of the normal positive sense of HPMood (Cummins, 2009, in press).

In this context, the validity of employing a structured clinical interview schedule would appear questionable since these have been developed within a categorical classificatory system. Specifically, and as outlined previously in this thesis, the DSM-IV-TR (American Psychiatric Association, 2000) and ICD-10 (WHO, 1992) adopt a disease approach to classifying psychological disturbance, centring on the notion of syndromes that form exclusive categories within a categorical framework. Here, diagnosis for disorder relies on the number of symptoms, or criteria met.

Within this categorical framework, a valid reference criterion for depression may take the form of a structured clinical interview capable of evaluating the diagnostic criteria for a depressive disorder as outlined within the DSM-IV-TR or ICD-10. In contrast, a dimensional approach to classifying psychological disturbance should take the form of continuous ratings along one or more dimensions (Mirowsky, 1994; Ruscio & Ruscio, 2000; Vrendernberg et al., 1993; Widiger, 1997), and symptom severity, as opposed to the number of symptoms, might constitute the criteria for ‘diagnosis’. As the self-report depression scales used in this study (BDI-II, DASS21-D and HADS-D) measure depressive symptoms along continuous ratings of severity, they may be more likely than diagnostic interviews to reflect the underlying structure of depression

as a dimensional entity (Beach & Amir, 2003; Franklin, Strong & Green, 2002; Ruscio & Ruscio, 2000). Further, if depression is dimensional, then SWB, as a measure of affective 'normality', should evidence utility as an indicator of depression.

In the current study, three ROC analyses were performed, each using one of the three self-report depression scales (BDI-II, DASS21-D and HADS-D) as the depression criterion. Thus, in each analysis the performance of two traditional self-report measures of depression and the same three SWB measures (HPMood, PWI-A and GLS) were together examined. In ROC analyses, test performance is evaluated by two methods. The first method comprises assessment of the relative magnitude of the areas under the ROC Curves (AUCs) for each test, with larger areas reflecting greater overall case-identification accuracy. This is followed by a statistical comparison of these AUC values. The second method, involves calculation of the maximal Youden Index (Youden, 1950), one of the most-widely used indices of optimal test performance, and subsequent examination of the sensitivity and specificity levels associated with this index for each test. Results pertaining to the overall accuracy of tests in differentiating depressed from normal cases (AUC values), and the optimal performance of tests (maximal Youden Index) will be discussed in turn.

Based on the relative magnitude of the AUC values, SWB measures achieve the following:

1. For BDI-II mild depression, HPMood provides the second best test performance, as it is superior to one of the depression scales (DASS21-D), the PWI-A and GLS.
2. For DASS21-D mild depression, PWI-A provides the best performance.
3. For HADS-D depression, HPMood provides the third best performance behind the two other depression scales.

However, pairwise comparisons revealed that no single instrument demonstrated a statistically significant superiority in detecting depression. While it is tempting to conclude that this finding is remarkable, in the context of

previous research (Low & Hubley, 2007; Meakin, 1992; Mulrow et al., 1995; Williams, Pignone, Ramirez & Stellato; 2002), failure to find a significant difference between case-finding instruments for depression is not unusual. In two previous comprehensive reviews of studies that have evaluated case-finding instruments for depression, authors examined the performance of sixteen case-finding instruments for depression against a diagnostic criterion standard in eighteen studies (Mulrow et al., 1995) and in 38 studies (Williams et al., 2002). In line with the current findings, no significant differences between any of the instruments were found.

It is noteworthy that the authors of the reviews on case-finding instruments for depression (Mulrow et al., 1995; Williams et al., 2002) fail to explain these negative results. It is, of course, possible that these instruments do not differ from each other. However, this seems unlikely given that the sensitivity and specificity of these instruments varies from 50 to 97% and 51 to 98% respectively. Another potential explanation is that inadequate power has contributed to these findings. This might also partly explain the failure to find significant differences between depression scales and SWB measures in the current study. Sample sizes required to ensure sufficient power for the statistical comparison of ROC Curves will now be examined.

Sample sizes for the statistical comparison of the areas under two ROC Curves.

Research studies that have undertaken ROC analyses generally refer to Hanley and McNeil's seminal paper (1982) on 'The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve'. In this publication, the authors provide formulae and tables to aid the design of research studies intending to undertake ROC analyses. One of these tables may be used to for planning sample sizes to ensure that one can statistically detect differences in the accuracy of tests. In this table, the number of cases required to provide a probability of 80%, 85%, or 95% of detecting various differences between Areas θ_1 and θ_2 under two ROC Curves using a one-sided test of significance with $p = 0.05$ is presented. AUC values are presented in increasing increments of 0.025 points. For Area θ_1, values range from 0.700 points to 0.950 points, and for Area θ_2, from

0.750 points to 0.975 points. Thus, these values may be seen to represent the range of interest for test accuracy, and allow researchers to estimate the sample size required to detect differences statistically between tests with a diagnostic (case-finding) accuracy of 70% to 97.5%. Importantly, the smaller the difference between areas, the larger the sample size required. Additionally, inspection of Hanley and McNeil's table indicates that fewer cases are required to detect a difference of 7.5% if it is a difference between 82.5% and 90% (N/group = 176) than if it is a difference between 72.5% and 80% (N/group = 267). Nevertheless, the number of cases required to detect these differences can be large.

Inspection of Hanley and MacNeil's (1982) table for sample sizes reveals that the number of cases required in the 'normal' and 'abnormal' group to ensure adequate statistical power are indeed considerable. For example, if the areas under two ROC Curves were 75.0% and 82.5%, one would need to plan on a sample with a minimum of 246 'normal' participants and 246 'abnormal' participants for each curve to have a reasonably high assurance (i.e. 80% probability) that the significance test would result in a statistically significant difference. For studies investigating the accuracy of tests in detecting depression, this corresponds to a sample of 246 normal participants and 246 depressed participants to test a difference of 7.5%.

The total number of cases (N=592) required to test a difference of 7.5% does not appear extraordinary. However, the number of depressed cases needed in the sample (N=246) is very large and may be unattainable in many research contexts. Specifically, previous research studies, such as those included in the reviews on case-finding instruments for depression described previously (Mulrow et al., 1995; Williams, 2002), have evaluated tests against a diagnostic criterion standard for Major Depression.

In one of these reviews, authors reported that the prevalence of Major Depression across eighteen studies using a variety of samples (Academic, Community, Screening Clinic, HMO) ranged from 4% to 17% (Mulrow et al., 1995). In these terms, the total sample required in the studies reviewed by Mulrow and colleagues to test a difference of 7.5% would range from 6,150 (Depression prevalence = 4%) to 1, 447 (Depression prevalence = 17%). The sample sizes of

the reviewed studies ranged from 31 to 809, and although sensitivity and specificity levels were reported for tests, AUC values were not. Nevertheless, estimating the difference between tests as 7.5% may be viewed as generous, as it may be reasonable to expect that differences would be smaller since the tests reviewed were validated against a criterion standard and were designed to measure the same construct.

Thus, in the context of Hanley and McNeil's recommendations, it is probable that Mulrow and colleagues (1995) failed to find any statistically significant differences between tests due to inadequate statistical power. It is therefore important to note that the time and resources required to acquire a sufficient number of people to ensure statistical sensitivity for a comparison between diagnostic tests may prove prohibitive for many researchers. Yet this is not the only difficulty associated with planning sample sizes using Hanley and McNeil's method.

In order to use Hanley and McNeil's (1982) table, one must first estimate the test accuracy (AUC value) of the two measures to be compared. These values may be estimated based on previous research findings, although it is likely that this will only provide a rough approximation of the sample size required as factors such as sample characteristics (for example, clinical *versus* general population) and the type of reference criterion used in analyses may vary across studies. Further, for studies investigating measures as novel tests of a particular phenomenon, previous estimates of AUC values may not exist. This is precisely the situation encountered in the current study. Nevertheless, while less than ideal, it is fortunate that Hanley and McNeil's table may be used post-hoc to verify negative results.

In the current study, results from pairwise comparisons of AUC values between independent ROC Curves indicated there were no statistically significant differences between tests in three separate analyses. Post-hoc assessment of these results using Hanley and McNeil's (1982) recommendations for sample sizes indicates that these negative results are likely the consequence of inadequate statistical power. Specifically, the sample size required per group to test the smallest difference found across all three analyses (HADS-D-HPMood for BDI-II

depression = 1.3%) is less than the smallest incremental difference (2.5%) listed in Hanley and McNeil's table, and as such does not correspond to any sample size. The sample size required to test the largest (PWI-A-GLS for DASS21-D = 8.1%) difference would be 267/group. Clearly, the number of positive cases in the three analyses (39, 44 and 50) performed in this study do not meet Hanley and McNeil's sample size recommendations, and therefore there is a real possibility that these negative results represent a Type II error. Notwithstanding this, meaningful qualitative conclusions may be made from ROC experiments that include approximately 100 observations (Metz, 1978). In these terms, inadequate power for the statistical comparison of AUC values does not prevent conclusions being drawn based on the relative magnitude of the AUC value of tests, and the performance of tests as evidenced by their sensitivity and specificity levels. The sensitivity and specificity of depression scales and measures of SWB will now be discussed.

The Optimal Performance of Tests

In the current study, three cut-scores were recommended for each of the depression scales and SWB measures in each of the three ROC analyses. The first cut-score represented maximal sensitivity with a specificity $\geq$ 60%, and was recommended for screening purposes. The second cut-score represented maximal specificity with a sensitivity $\geq$ 60%, and was recommended for research studies.

The parameters for these two cut-scores were study-specific and provide an indication of test performance potential for screening (priority = sensitivity) and research (priority = specificity) purposes. However, each of these separate scores levies a substantial sacrifice in specificity for screening purposes, and in sensitivity for research purposes.

The third cut-score recommended for each of the measures reflects the maximal Youden Index (Youden, 1950). As previously described in the results section, this index is one of the most widely used summary measures of ROC Curves and represents the maximum potential effectiveness of a test (Ruopp, Perkins, Whitcomb & Schisterman, 2008). It is calculated as (sensitivity + specificity-1) for all possible cut-scores on a test. The cut-point that corresponds

to the maximal Youden Index may be referred to as the optimal cut-point, as it is the cut-point that optimises the tests ability to discriminate cases from non-cases when equal weight is given to sensitivity and specificity (Fluss, Faraggi & Reiser, 2005; Youden, 1950; Reiser, 2000; Ruopp, Perkins, Whitcomb & Schisterman, 2008). Additionally, if the level of sensitivity that corresponds to the optimal cut-point, and therefore to the maximal Youden Index, is higher than the associated specificity level, then the optimal performance of the test would indicate that this test may have greater utility as a screening tool. However, if the level of specificity that corresponds to the optimal cut-point of the test were higher than the associated sensitivity level, then the optimal performance of the test would indicate that this test might be classified as more effective for research purposes.

Based on the relative sensitivity and specificity levels associated with the maximal Youden Index for all tests, SWB measures achieve the following:

For BDI-II mild depression, HPMood demonstrates the highest maximal Youden Index, and of the three scales with higher sensitivity than specificity, HPMood is the most effective test for screening purposes, being superior to one other depression scale, (DASS21-D and GLS).

For DASS21-D mild depression, the PWI-A demonstrates the highest maximal Youden Index, and of the three scales with higher sensitivity than specificity, the PWI-A is the most effective test for screening, being superior to one other depression scale (BDI-II and HPMood).

For HADS-D Depression, HPMood demonstrates the third highest maximal Youden Index, and of the three scales with higher sensitivity than specificity, is the third most effective screening tool, being inferior to the two other depression scales.

In combination, these findings indicate that measures of SWB, as HPMood and the PWI-A, demonstrate the greatest effectiveness in detecting mild depression of all the tests used. Obviously, these findings require replication, and further research is required to establish cut-scores for SWB measures to identify increasing levels of depression severity. Nevertheless, these findings highlight the possibility that SWB measures may be useful for depression screening in primary

care settings, and may have potential as measures of depression risk in public health initiatives (e.g., Beyond Blue 2000) focussed on the early detection of depressive symptoms, and prevention of depressive disorders.

These findings are also consistent with previous research that has found the underlying structure of depression to be dimensional (Beach & Amir, 2003; Franklin, Strong & Green, 2002; Ruscio & Ruscio, 2000, 2002). In these terms, and as argued within this thesis, depression may be considered on a continuum with affective normality, which may be best described in terms of normative parameters for HPMood and SWB within the Cummins' theory of SWB homeostasis.

In conclusion, if depression is a dimensional entity, the continued reliance on diagnostic interviews, developed within a taxonic (categorical) framework as the 'gold standard' reference criterion in criterion validity research, appears ill advised. However, it is important to note that while there is a growing body of empirical evidence to support the notion of a continuous (dimensional) structure for most, if not all, mental disorders (Beach & Amir, 2003; Eyesenck et al., 1983; Franklin et al., 2002; Gunderson, Links & Reich, 1991; Mirowsky, 1994; Persons, 1986; Ruscio & Ruscio 2000, 2002; Widiger, 1997) the taxonic *versus* continuity question has been debated for more than one hundred years and to date, remains unresolved.

In these terms, it is imperative that researchers ensure that there is congruence between the reference criterion (self-report scales using continuous ratings and structured clinical interviews) and the conceptual frame of the research question (dimensional and categorical) within their study. In the context of the current report, there is such congruence between the conceptual framework (dimensional) and the reference criteria (self-report scales using continuous ratings). Thus, the results may be seen as valid and interpretable, and as such, are remarkable.

CHAPTER 11: OVERVIEW AND CONCLUSIONS

This thesis explores the relationship between SWB and depression. It investigates the theoretical proposition that depression may be conceptualised as a loss of the normal positive sense of HPMood subsequent to the failure of the homeostasis (Cummins, 2009, in press). While it was argued in this thesis that this suggestion is meritorious, the validity of this theoretical proposition, and therefore the investigations undertaken herein, relies on the assumption that SWB and depression are valid in terms of theoretical structure, composition, definition and measurement.

A review of the literature related to these conceptual characteristics for SWB and depression confirmed that many of these issues remain contentious, retarding our understanding of the fundamental nature of these constructs that underpins the interrogation of more complex questions. One of these questions relates to the mechanisms underlying the maintenance of SWB, a presumably key area in the context of Cummins' (2009, in press) abovementioned theoretical proposition and escalating rates of depression globally (WHO, 2001). In these terms, it would appear critical that researchers plan and design studies carefully, define terms, and select measures that are consistent with the conceptual frame of their research question and the construct being examined.

This research addresses these conceptual issues by investigating SWB and depression from within the same framework, that is, from a dimensional understanding of affective experience. Here, mental disorders such as depression are commonly viewed as existing on a continuum with normality (Eyesenck et al; Widiger, 1997). Major Depression represents the end-point along a continuum of depressive symptomatology, as opposed to it being a structurally distinct entity (Ruscio & Ruscio, 2000). Moreover, a dimensional framework is consistent with the proposition that depression represents the loss of the normal positive sense of HPMood and with the understanding and measurement of affect within SWB homeostasis theory (Cummins & Nistico, 2002). This theory posits that SWB is actively managed by a system of psychological devices that have evolved for this purpose, and that this management is actually directed at the protection of HPMood, which is the major component of SWB. Further, in this dimensional

context, it may be more likely that self-report depression scales that measure depressive symptoms along continuous ratings of severity, represent more valid reference criteria for depression than structured diagnostic interviews. This is because the latter have been developed within a categorical taxonomy for mental disorders. Consequently, investigations in this thesis used self-report depression scales to measure depression.

The investigations in this thesis comprise two linked studies. The first study tested a number of the theoretical predictions and diagnostic approximations made within SWB homeostasis theory (Cummins & Nistico, 2002). These involved testing the relationship between SWB and depression as a means to verify homeostasis as the proposed mechanism of SWB control and assess the viability of the proposition that depression represents a loss of normal levels of SWB. In this study, SWB was measured using the PWI-A (International Wellbeing Group, 2006), and depression was measured using DASS21-D (Lovibond & Lovibond, 1995).

The overall findings provided general support for homeostasis as the proposed mechanism of SWB control, and gave rise to a number of additional propositions regarding the relationship between SWB and depression. Namely, if depression represents the loss of HPMood (Cummins, 2009), and the underlying structure of depression is dimensional, as found previously (Beach & Amir, 2003; Franklin et al., 2002; Ruscio & Ruscio 2000, 2002), then SWB as a measure of affective 'normality' should evidence utility as an indicator of depression. This latter proposition was investigated in the second study using three measures of SWB and three traditional self-report depression scales. For SWB, these measures were HPMood (Cummins 2009), the PWI-A (International Wellbeing Group, 2006), and GLS (International Wellbeing Group, 2006). For depression, these measures were the BDI-II (Beck et al., 1996), the DASS21-D (Lovibond & Lovibond, 1995), and the HADS-D (Zigmond & Snaith, 1983).

The second study produced unquestionably the most interesting finding within this thesis; that some SWB measures performed as well as depression measures in identifying cases of depression, and performed better than depression scales to screen for depression. An overview of these findings and their

implications will now be presented. Methodological limitations of this research and recommendations for future research are presented in the context of these findings throughout this chapter.

Homeostasis as the Mechanism of SWB Control

In an attempt to account for the determined positivity and stability of SWB, Cummins and colleagues have described the regulation of SWB as being consistent with a homeostatic mechanism. Describing regulation in these terms permits very clear predictions regarding the behaviour of SWB, as well as the relationship between SWB and other variables. These include: (1) it must be highly stable; (2) there must be a threshold value which is being defended by homeostatic processes; (3) there must be evidence that, as this value is approached, the system works harder than normal to retain control evidenced by a 'plateau effect' in SWB levels despite increasing strength of challenge; and (4) once the threshold value is exceeded, there must be evidence that homeostasis has failed and is no longer controlling the level of SWB. This should be confirmed by a dramatic fall in levels of SWB. In other words, in addition to stability, SWB must resist change and demonstrate a curvilinear relationship with variables that reflect a challenge to homeostasis that is consistent with Figure 3. These characteristics were tested in this thesis and results are now presented.

Subjective Wellbeing is Normally Positive

With only 7.03% of the total sample (N=3,700) from Study 1, and 5.72% of the total sample (N=559) from Study 2, reporting SWB levels in the negative sector of the response scale, these results support the notion that SWB is naturally positive. These findings are consistent with the results from all publications investigating SWB in general population studies from Western nations, and confirmed by results from the 21 surveys of the Australian Unity Wellbeing Project (Cummins et al., 2003) that have tracked the SWB of Australian adults over the last nine-year period (2001-2009). In combination, these results indicate that it is normal for people to feel good about themselves, and this is in accordance with SWB providing the motivation for behaviour across the lifespan.

Subjective Wellbeing is Highly Stable

The fundamental principle of any homeostatic system is stability. A review of the literature confirmed that SWB is stable, and therefore predictability, over short (Bacharowski & Braaten, 1994; Diener & Larsen, 1984; Larsen & Diener, 1987; Pavot & Diener, 1993) and long (Baur & Oken, 1983; Bowling et al., 1996; Costa & McCrae, 1989; Eid & Diener, 2004; Diener & Larsen, 1984; Heady & Wearing, 1989; Watson & Walker) periods respectively, and in Western populations (Cummins, 1995, 1998).

The remarkable stability of SWB is further highlighted by the extraordinarily narrow Australian adult normative ranges for SWB derived from AUWI surveys over a nine-year period. For the PWI-A the normative range is 73.43 to 76.43 points, and for GLS (satisfaction with life as a whole) the normative range is 75.2 to 79.1 points (Report 21.0, Cummins et al., 2009). Notably, in this thesis, the mean score for the PWI-A (73.4) in Study 1, and the mean scores for the PWI-A (74.66) and GLS (75.62) were consistent with these normative statistics. Additionally, in Study 2 the mean score for HPMood, the major component of SWB, is also consistent with the normative ranges for both the PWI-A and GLS. In combination, these results attest to the stability of SWB, and the reliability of SWB measures used in this thesis.

These results provide support for the notion that SWB is controlled by a homeostatic system. Further, in line with a dimensional understanding of affective experience, it is feasible that the normative ranges for SWB found in Australia, may be used to define the parameters for mental health in this country. Nevertheless, the investigation of depression in the context of differing SWB score ranges produced some apparent inconsistencies with theory.

Apparent inconsistencies with SWB homeostasis theory.

The results from Study 1 and Study 2 indicate that the prediction of a strong inverse relationship between SWB and depression was confirmed. This is consistent with findings from numerous other studies (Cheung & Badgely, 1998; Cook & Cummins, 2003; Davern & Cummins, 2004; Heady et al., 1993; Hong & Giannakopoulos, 1994; Lewis et al., 1999; Lewisohn et al., 1991; Simpson et al.,

1996) and theory. Moreover, in accordance with expectation from homeostasis theory, results demonstrated that the risk for depression increases substantially at a level of SWB <50%SM. However, the proportion of people (23.9%) reporting depression in conjunction with normative range (70-80 points) SWB far exceeds expectation. Additionally, the proportion reporting depression together with SWB less than 50 points (87.8%) is far less than what is theoretically predicted. However, it is important to note that such an interpretation relies on the assumption that the DASS is valid in terms of the presumed categories of depression. Notwithstanding this, these results are not easily explained within homeostasis theory as it is currently articulated. It is possible that a more detailed analysis of the relationship between SWB and DASS depression may yield further insights into whether this is a flawed prediction from homeostasis theory or whether it has some other explanation. This analysis was undertaken in the second part of Study 1 where evidence for the presence of a curvilinear relationship between SWB and depression was examined.

Subjective Wellbeing Resists Change

Cummins and Nistico (2002) have proposed that the critical evidence for the operation of homeostatic processes will be demonstration of a curvilinear relationship between SWB levels and increasing levels of challenge, consistent with Figure 3. Specifically, and as described previously in this thesis, these researchers have suggested that as the strength of challenge to homeostasis progressively increases, mean PWI-A scores evidence a smooth downward trajectory until the lower threshold for homeostasis is reached at about 70 points. The explanation they offer is that, at any point in time, the exact location of SWB within the set-point range is a probability statement determined by the balance of challenge and homeostatic support. As progressively higher levels of challenge are applied to the homeostatic system, the probability increases that SWB will inhabit the lower reaches of the normative range. As this happens, SWB increasingly becomes bonded to the lower threshold, where it appears to hold steady, thereby demonstrating a plateau effect. However, as the level of challenge continues to strengthen, the homeostatic defence system is eventually overwhelmed and fails, evidenced by a marked drop in SWB. Further, these

researchers claim that evidence for this exists in the finding of a curvilinear relationship between SWB and depression from an earlier investigation conducted by Davern (2003, unpublished).

Consistent with the principles of homeostasis, and earlier research conducted by Davern (2003, unpublished), the results from Study 1 indicate a plateau in SWB scores. Specifically, the rate at which SWB decreases attenuates across four contiguous depression score groups (DASS21-D scores 15.1-27.0), where SWB scores differ sequentially by < 2 points, falling a total of 4.1 points. This is contrasted with the amount of change in SWB for the four immediately higher (10.1 points) and lower score groups (13.9 points). Notably, the largest decrease in SWB levels (7.7 points) occurs immediately after the plateau.

These results are consistent with the curvilinear relationship between SWB and increasing levels of challenge outlined above. Additionally, these findings furnish support for theoretical predictions, are in accordance with homeostasis as the proposed mechanism of SWB control, and are consistent with the proposition that depression represents the loss of the normal positive sense of HPMood.

Perhaps the most interesting question raised by these results, is why plotting the PWI-A against the DASS21-D shows this trend towards a plateau effect. This question may be elaborated within the context of SWB homeostasis theory. Specifically, the DASS items measure the degree of negative affect (down-hearted and blue), life being meaningless, low personal self-worth, and hopelessness. In other words, even quite strong negative feelings about the self can co-exist with normal or even high levels of SWB. Other results from Study 1 support this proposition, with an unexpectedly high proportion (23.9%) of the people reporting a SWB of 70-80 points also reporting some level of DASS depression. Thus, there is a degree of disconnection between negative and positive feelings about the self as long as homeostasis is functional. It is proposed that this is highly adaptive in allowing negative feelings to be acknowledged while also maintaining normal levels of SWB. However, once the level of challenge becomes overwhelming, positive feelings about the self evaporate and it is possible that true depression sets in. Thus, the demonstration of a trend towards a plateau in SWB scores provides support for the operation of homeostatic

processes to maintain normal positive levels of HPMood, and this is consistent with theory.

Notwithstanding this trend, analyses in Study 1 failed to verify statistically a significant plateau in SWB levels, although it is likely that has a methodological explanation. Specifically, it is likely that the depression scale of the DASS functions as an indicator variable, rather than a causal variable. Indicator and casual (challenge) variables differ in the relationship they have with the construct being measured. Whereas casual variables affect the level of SWB, the changing level of SWB itself is an indicator variable (Fayers & Hand, 1997; Fayers, 2002). As the DASS measures depression through parameters that are essentially the inverse of positive affect (e.g., anhedonia), depression is not causal. Depression measured in this way is a measure of outcome, and so is an indicator variable, just like SWB. Moreover, as the DASS21-D reflects the psychological consequences of the loss of the normal levels of positive mood (e.g., hopelessness), then such measures of outcome may be virtually indistinguishable from SWB. In these terms, failure to confirm a statistically significant plateau in SWB levels appears reasonable. Yet, the visual emergence of the plateau is undeniable.

Closer inspection of the pattern of change in SWB and depression items across increasing depression scores suggested that the trend towards a plateau, and therefore the appearance of a curvilinear relationship between SWB and depression, can be explained by the casual-indicator variable distinction. Specifically, the inclusion of two more cognitive symptoms as items (self-deprecation and devaluation of life), in the predominantly affective seven-item DASS21-D scale, likely accounts for the apparent trend towards a curvilinear relationship between SWB and depression. This is because these two more cognitive items are less driven by HPMood, and therefore are more like casual variables than the affect items. Results supported this conclusion revealing that the items self-deprecation and devaluation of life had mean scores least like the reciprocal of the PWI-A, and had trajectories across increasing depression scores that were significantly different from the PWI-A. These results are consistent with the idea that the items measuring self-deprecation and devaluation of life are more cognitive, are less driven by HPMood, and therefore function more like causal

variables, and as reflecting measures of challenge. Thus, the inclusion of two items that reflect a challenge to homeostasis may explain the emergence of a plateau in SWB levels, and the appearance of a curvilinear relationship between SWB and depression, consistent with theory.

In combination, results from Study 1 and Study 2 provide a range of evidence that are consistent with homeostasis as the proposed mechanism of SWB control. Specifically, results in this thesis indicate that SWB is highly stable, and therefore predictable. Further, results indicate that SWB resists change in the context of variables that reflect a challenge to homeostasis, evidenced by a plateau in SWB levels despite increasing levels of challenge. Additionally, results demonstrated that SWB levels fell markedly directly after the 'plateau effect', supporting the notion that homeostasis has failed, and is no longer in control of SWB, and this is consistent with Figure 3. Moreover, these results strengthen the argument that depression represents the outcome of the loss the normal sense of HPMood subsequent to the failure of homeostasis, and this is in line with a dimensional understanding of affective experience.

These conclusions raised the possibility that measures of SWB may represent indices of depression and this idea led to the generation of a number of additional hypotheses regarding the relationship between SWB and depression, and particularly, the utility of SWB measures as novel measures of depression. These formed the basis for Study 2, and were tested in this study.

The Comparative Validity of SWB and Depression

The major aim of the second study was to investigate the comparative validity of three widely used depression scales (BDI-II, DASS21-D and HADS-D) and three measures of SWB (HPMood, PWI-A and GLS) as putative, novel measures of depression. Clearly, the validity of these investigations relies on the reliability and validity of SWB and depression measures.

The manner in which SWB and depression scales operationalise their respective constructs will be outlined. Following this, the utility of SWB scales as measures of depression is presented.

SWB Measures Essentially Measure the Same Construct

Extending the findings of Study 1, results from Study 2 indicated that all three measures of SWB (HPMood, PWI-A and GLS) approximate each other. Their mean scores varied less than one percentage point and did not differ significantly. These results are consistent with theory, and earlier research which has found that HPMood is the major component of SWB (Davern & Cummins 2004; Tomyn & Cummins, 2008) and that the PWI-A represents a first-level deconstruction of GLS (International Wellbeing Group, 2006). Moreover, these findings indicate that SWB measures are essentially measuring the same construct, and further attest to their reliability and stability.

Depression Scales Differ in the Conceptualisation and Measurement of Depression

In contrast to the similarities found between measures of SWB, the mean scores for the three traditional self-report depression scales differed significantly. Specifically, there were no significant differences between the DASS21-D and HADS-D, although both of these scales had mean scale scores that were significantly higher (+2.38; +2.73 points) than the BDI-II mean. These results are consistent with differences in scale content (DASS21-D, HADS-D, psychological symptoms; BDI-II, psychobiological symptoms consistent with DSM-IV diagnostic criteria for MD) and reporting periods for symptoms (DASS21-D, HADS-D, one week; BDI-II, two weeks) across the three depression scales. More particularly, the finding of significantly higher means for the DASS21-D and HADS-D is likely the consequence of differences in scale content being attenuated by differences in reporting periods. As the BDI-II contains items that may be related to physical illness as well as depression, the BDI-II may measure more than one condition. As such, it would be expected that the BDI-II would yield higher total scores, and therefore demonstrate a higher mean score, than the DASS21-D and HADS-D, which measure only those symptoms that are unique to depression. However, this is not correct.

The BDI-II was designed to be consistent with a medical definition of depression, with scale items and the reporting period for symptoms in line with

DSM-IV diagnostic criteria for an episode of Major Depression. Importantly, the number of symptoms and the duration of symptoms is the key to differentiating Major Depression (five of nine DSM-IV diagnostic criteria, must include anhedonia or depressed mood over last two weeks) from Dysthymic Disorder and subthreshold depression (< 5 symptoms, < 2 weeks) (Ingram & Siegle). In these terms, the BDI-II represents a measure of clinical or Major Depression, whereas the DASS21-D and HADS-D represent measures of subthreshold depression. Due to the relaxation of inclusion criteria for subthreshold depression, its frequency is manifold higher in general population samples than diagnosable depressive disorders (Judd et al., 1994; Judd et al. 1997). In these terms, as self-report depression scales measure the frequency/severity of depressive symptoms, the DASS21-D and HADS-D should demonstrate higher standardised total scores, and higher mean scores than the BDI-II as a measure of clinical depression. This was precisely the finding in this thesis. Other results from Study 2 support this conclusion, with the finding of a significantly higher frequency of cases classified as normal according to the BDI-II (84.6%) relative to the DASS21-D (77.1%).

In conclusion, the scale content of depression scales, and the reporting period for symptoms, reflects the framework within which depression is conceptualised (BDI-II, categorical; DASS21-D & HADS-D, dimensional) and defined (BDI-II, medical; DASS21-D & HADS-D, psychological), and therefore what is being measured (Depressive Disorders *versus* subthreshold depression). As such, these differences should be considered and used to guide researchers and clinicians in selecting the most appropriate instrument for their specific purposes. Notwithstanding these definitional implications, manipulation of the reporting period for symptoms has implications for the measurement of affect in general and for the relationship between SWB and depression within homeostasis theory.

The Implications of Reporting Periods for the Measurement of Affect

The one-week reporting period for depressive symptoms aims to elicit respondent reports of states as opposed to reports based on traits (how the respondent characteristically feels or experiences) or Momentary Emotions (how the respondent feels right now) (Lovibond & Lovibond, 1995). Differentiating these three types of respondent reports is consistent with the three levels of affect

within the affective framework of SWB homeostasis theory. While the depression scales used in this thesis cast items in the state form (over the past week or two weeks), SWB measures generally cast items in the trait form: HPMood – how the content, alert, and happy the respondent generally feels each day; PWI-A – how the respondent generally feels about different aspects of their life; and GLS – how the respondent generally feels about their life as a whole.

As SWB and depression scales aim to elicit responses according to different levels of affect (trait *versus* state), it is possible that there is a level of incongruence in the measurement of affect between the two types of scales. This should not pose a problem when the homeostatic system is functioning normally. This is because at these times, Current Mood (state) should approximate the set-point for HPMood (trait). Therefore, under normal operating conditions, only two levels of affect are experienced or felt by the individual, and for the purpose of measurement, Current Mood (state) and HPMood (trait) will not differ despite the use of different reporting period for affect items.

In these terms, if depression of a certain level of severity presents in the context of normal SWB, then this may provide evidence that the individual's set-point range for HPMood has been breached and homeostasis has failed. This is because in the event of strong, prolonged negative challenge, homeostasis eventually fails, and the perception of HPMood falls out of the individual's genetically determined set-point range. At this point, an individual loses contact with HPMood, Current Mood reflects the challenging agent, and the individual's dominant affective experience will be negative. Here the loss of positive HPMood and the consequent separation of Current Mood from HPMood is the essence of depression. In these terms, HPMood and Current Mood (depression) may be differentially measured, although they will continue to evidence a strong inverse linear relationship as found in this thesis and in previous research.

Nevertheless, further research is required to investigate the impact of manipulating the reporting period when measuring affect. More explicitly, research is required to investigate whether altering the reporting period for SWB measures to reflect a one- to two-week period may increase their sensitivity and specificity as indices of depression and further clarify the relationship between

SWB and depression. Notwithstanding this, results from three ROC analyses in this thesis indicate that some measures of SWB have greater utility in screening for depression than some of the most widely used traditional self-report depression scales specifically designed for this purpose.

SWB has Utility as a Measure of Depression

The accuracy of the three SWB measures and three self-report depression scales to differentiate cases of depression from normal cases are investigated in three ROC analyses. Each analysis used one of the self-report depression scales as the reference criteria for depression. Therefore, in each analysis the performance of three SWB measures and two depression scales were examined.

A depression rating of mild depression defined a positive case in each ROC analysis. A level of mild depression was selected based on the finding in this thesis that the level of depression associated with the failure of homeostasis (SWB < 70 points) differs depending on the response format of the depression scale (4-point = mild depression; 11-point = moderate depression). As depression scales in ROC analyses used their original 4-point response scales, a depression rating of mild is the most appropriate parameter to define a positive case of depression in this thesis. For future research studies conducting similar investigations using depression scales with 11-point response scales, a depression rating of moderate should be used to define a positive case.

The relative performance of tests was evaluated using two methods. The first was based on the relative magnitude of the AUC for each scale. Statistical comparison of AUC values indicated that there were no significant differences between the overall performance of tests. This finding is consistent with virtually all publications reporting ROC results on case-finding instruments for depression. Post-hoc examination of these results using Hanley and McNeil's formulae for estimating sample sizes required to statistically test differences between ROC Curves revealed that these negative results are most likely the consequence of inadequate statistical power. This comprises a limitation of this research. Nevertheless, it is important to note that sample sizes recommended by Hanley and McNeil are extremely large. Consequently, the time and resources required to

attain a sufficient number of people reporting depression to ensure statistical sensitivity for a comparison between diagnostic tests proved prohibitive in this research. Further, it is likely that this also proved prohibitive in other research studies, and may explain the failure to demonstrate statistically the superiority of one instrument over another in virtually all publications on case-finding instruments for depression.

In the context of rapidly escalating rates of depression globally, it is imperative that physicians and researchers are able to select and utilise the most accurate case-finding instrument to detect the early signs of depression to take preventative action, or refer appropriately. In these terms, considerable efforts should be directed towards the formulation of a more parsimonious and pragmatic way of statistically comparing the areas under independent ROC Curves, so that future ROC research is able to demonstrate the relative superiority of case-finding instruments for depression definitively. Presently, evaluation of case-finding instruments is a more subjective process, requiring the examination of cut-scores and their associated sensitivity and specificity pair. This comprises the second method for assessing test performance in this thesis.

Three cut-scores were presented for each scale in each of the three analyses. The first cut-score represented maximal sensitivity with a specificity $\geq$ 60%, and was recommended for screening purposes. The second cut-score represented maximal specificity with a sensitivity $\geq$ 60%, and was recommended for research studies. The parameters for these two cut-scores were study-specific and provide an indication of test performance potential for screening (priority = sensitivity) and research (priority = specificity) purposes. However, each of these separate scores levies a substantial sacrifice in specificity for screening purposes, and in sensitivity for research purposes. Consequently, cut-scores generated according to these parameters are recommended for use with the caveat that clinicians and researchers carefully consider the possible implications of such a trade-off between sensitivity and specificity when selecting instruments. Notwithstanding this, a third cut-score and its associated sensitivity and specificity pair presented for each test provide another way to ascertain whether instruments perform best as clinical or research tools.

The third cut-score is generated according to one of the most widely used summary measures of ROC Curves, known as the maximal Youden Index. This index represents the maximum potential effectiveness of a test when equal weight is given to sensitivity and specificity. It corresponds to an optimal cut-point, with an associated sensitivity and specificity pair. In these terms, if the sensitivity associated with the maximal Youden Index is higher than the specificity, then the optimal test performance indicates that the instrument is most effective as a screening tool, whereas if specificity is higher than sensitivity, then the test is more appropriate for research purposes.

Results indicated that HPMood and the PWI-A demonstrate the greatest effectiveness in detecting mild depression of all the tests used for BDI-II and DASS21-D depression respectively, and comprise the most effective instruments for use as screening tools according to these same depression criteria. These findings are remarkable. Not only do measures of SWB outperform some of the most widely used depression scales specifically designed to detect cases of depression, but they do so in the context of two differing depression criteria that represent two differing conceptualisations and formulations of depression (DASS21-D, dimensional, psychological symptoms; BDI-II, medical, psychobiological symptoms).

Obviously, these findings require replication. Additionally, further research is required to establish cut-scores for SWB measures to identify increasing levels of depression severity. Nevertheless, these results suggest HPMood and the PWI-A may have greater utility than traditional self-report depression scales in identifying the early signs of depression. In these terms, these findings highlight the possibility that SWB measures may be useful for depression screening in primary care settings, and they may have great potential as measures of depression risk in public health initiatives (e.g., Beyond Blue, 2000) focussed on the early detection of depressive symptoms, and prevention of depressive disorders. Further, as the PWI-A comprises eight items of satisfaction with different life domains, this scale is not only capable of detecting depression as found in this thesis, but is able to provide information regarding the areas of an individual's life that are associated with the depression. In this way, unlike extant

depression that mostly measure levels of negative affect and negative cognitions, the PWI-A may be used to highlight life areas that should be targeted by interventions.

Overall, these results directly link this research to the dimensional-categorical debate, and are consistent with previous research that has found the underlying structure of depression to be dimensional (Beach & Amir, 2003; Franklin, Strong & Green, 2002; Ruscio & Ruscio, 2000, 2002). In these terms, and as argued within this thesis, depression may be considered on a continuum with affective normality, which may be best described in terms of normative parameters for HPMood and SWB within the theory of SWB homeostasis.

It is undeniable that the manner in which we define, conceptualise, operationalise and, importantly, measure constructs affects what we study, how we design and conduct research and, inevitably, the interpretability of our results and the conclusions we draw. This thesis consistently highlights these conceptual characteristics as a means to draw attention to their fundamental importance, and the need for researchers and clinicians to consider these elements carefully at each point in the research and assessment process respectively to ensure results are meaningful.

If depression is a dimensional entity, the continued reliance on diagnostic interviews, developed within a taxonic (categorical) framework as the ‘gold standard’ reference criterion in criterion validity research, appears ill advised. However, it is important to note that while there is a growing body of empirical evidence to support the notion of a continuous (dimensional) structure for most, if not all, mental disorders (Beach & Amir, 2003; Eyesenck et al., 1983; Franklin et al., 2002; Gunderson, Links & Reich, 1991; Mirowsky, 1994; Persons, 1986; Ruscio & Ruscio 2000, 2002; Widiger, 1997) the taxonic *versus* continuity question has been debated for more than one hundred years and to date, remains unresolved.

For future research on SWB and depression, it is imperative that researchers ensure that there is congruence between the reference criterion (self-report scales using continuous ratings and structured clinical interviews) and the

conceptual frame of the research question (dimensional and categorical) within their study. In the context of the current report, there is such congruence between the conceptual framework (dimensional) and the reference criteria (self-report scales using continuous ratings). Thus, the results herein may be seen as valid and interpretable, and as such, are remarkable.

References

Abbey, A. & Andrews, F.M. (1984). Modeling the psychological determinants of life quality. *Social Indicators Research, 16,* 1-34.

Abdi, H. (2007). Bonferroni and Sidak corrections for multiple comparisons. In N.J. Salkind (Ed.), *Encyclopaedia of Measurement and Statistics* (pp. 540-542). Thousand Oaks, CA: Sage.

Abou-Saleh, M.T., Ghubash, R. & Daradkeh, T.K. (2001). Al Ain community psychiatric survey. I. Prevalence and socio-demographic correlates. *Social Psychiatry and Psychiatric Epidemiology, 36,* 20-28.

Akiksal, H.S., Judd, L.L., Gillin, J.C. & Lemmi, H. (1997). Subthreshold depressions: Clinical and polysomnographic validation of dysthmic, residual and masked forms. *Journal of Affective Disorders, 45*, 53-63.

Akin, A. & Cetin, B. (2007). The Depression Anxiety and Stress Scale (DASS): The study of reliability and validity. *Educational Sciences: Theory and Practice*, *7*(1), 260-268.

Almeida-Filho, N., Mari, J.D.J., Coutinho, E., France, J.F., Fernandes, J., Andreoli, S.B. et al. (1997). Brazilian multicentric study of psychiatric morbidity. *British Journal of Psychiatry*, *171*, 525-529.

Al Musawi, N. (2001). Psychometric properties of the Beck Depression Inventory-II with university students in Bahrain. *Journal of Personality Assessment, 77*, 568-579.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders – Fourth edition.* Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders – Fourth edition – Text revision*. Washington, DC: American Psychiatric Association.

American Psychological Association. (2007). *APA dictionary of psychology*. Washington, DC: American Psychological Association.

Andrade, L., Walters, E.E., Gentil, V. & Laurenti, R. (2002). Prevalence of ICD-10 mental disorders in a catchment area in the city of Sao Paulo, Brazil. *Social Psychiatry and Psychiatric Epidemiology, 37*, 316-325.

Andreasen, N.J.C. & Norris, A.S. (1972). Long-term adjustment and adaptation mechanisms in severely burned adults. *The Journal of Nervous and Mental Disease, 154*, 352-362.

Andrews, G., Henderson, S. & Hall, W. (2001). Prevalence, comorbidity, disability and service utilisation. Overview of the Australian National Mental Health Survey. *British Journal of Psychiatry, 178*, 145-153.

Andrews, G. & Peters, L. (1998). The psychometric properties of the Composite International Diagnostic Interview. *Social Psychiatry and Psychiatric Epidemiology, 33*(2), 80-88.

Andrews, G., Peters, L., Guzman, A.M. & Bird, H. (1995). A comparison of two structured diagnostic interviews: CIDI and SCAN. *Australian and New Zealand Journal of Psychiatry, 29*, 124-132.

Andrews, F.M. & Withey, S.B. (1976). *Social indicators of well-being: American's perceptions of life quality*. New York: Plenum Press.

Angst, J. & Merikangas, K. (1997). The depressive spectrum: Diagnostic classification and course. *Journal of Affective Disorders, 45*(1-2), 31-40.

Annick, M. (2002). The Chilean health system: 20 years of reforms. *Salud Publica Mex, 44*, 60-68

Antony, M.M., Bieling, P.J., Cox, B.J., Enns, M.W. & Swinson, R.P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment, 10*(2), 176-181.

Araya, R., Rojas, G., Fritsch, R., Acuna, J. & Lewis, G. (2001). Common mental disorders in Santiago, Chile: Prevalence and socio-demographic correlates. *The British Journal of Psychiatry*, *178*(3), 228-233.

Araya, R., Wynn, R., Leonard, R. & Lewis, G. (1994). Psychiatric morbidity in primary health care in Santiago, Chile. Preliminary findings. *British Journal of Psychiatry*, *165*, 530-533.

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

Archer, R.P., Maruish, M., Imhof, E.A. & Piotrowski, C. (1991). Psychological test usage with adolescent clients: 1990 survey findings. *Professional Psychology: Research and Practice*, *22*(3), 247-252.

Aspinwall, L.G. (1998). Rethinking the role of positive affect in self-regulation. *Motivation and Emotion*, *22*, 1-32.

Bach, J.R. & Tilton, M.C. (1994). Life satisfaction and well-being measures in ventilator assisted individuals with traumatic tetraplegia. *Archives of Physical Medicine and Rehabilitation*, *75*, 626-632.

Bachorowski, J.A. & Braaten, E.B. (1994). Emotional intensity: Measurement and theoretical implications. *Personality and Individual Differences*, *17*, 191-199.

Bartlett, M.S. (1954). A note on multiplying factors for chi square approximations. *Journal of the Royal Statistical Society*, *16*(Series B), 296-298.

Baur, P.A. & Okun, M.A. (1983). Stability of life satisfaction in late life. *The Gerontologist*, *23*, 261-265.

Beach, S.R.H. & Amir, N. (2003). Is depression taxonic, dimensional, or both? *Journal of Abnormal Psychology*, *112*(2), 228-236.

Bebbington, P. (2004). The classification and epidemiology of unipolar depression. In M. Power (Ed.), *Mood disorders: A handbook of science and practice* (pp. 3-29). Chichester: John Wiley & Sons.

Beck, A.T. (1967). *Depression: Causes and treatment*. Philadelphia: University of Pennsylvania Press.

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

Beck, A.T. & Steer, R.A. (1990). *Manual for the Beck Anxiety Inventory*. San Anotonio, TX: The Psychological Corporation.

Beck, A.T. & Steer, R.A. (1993). *Beck Depression Inventory manual*. San Antonio, TX: Harcourt Brace & Company.

Beck, A.T., Steer, R.A. & Brown, G.K. (1996). *BDI-II Manual*. San Antonio, TX: The Psychological Corporation.

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

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

Beckham, E.E., Leber, W.R. & Youll, L.K. (1995). The diagnostic classification of depression. In E.E. Beckham & W.R. Leber (Eds.), *Handbook of depression: Treatment, assessment and research* (2nd ed.) (pp. 36-60). New York: Guilford.

Bonciatto, S.C., Dew, M.A., Zaratiegui, R., Lorenzo, L. & Pecina, P. (2001). Adult outpatients with depression: Worse quality of life than in other chronic medical diseases in Argentina. *Social Science and Medicine*, *52*, 911-919.

Bjelland, I., Dahl, A.A., Haug, T.T. & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale: An updated literature review. *Journal of Psychosomatic Research*, *52*(2), 69-77.

Bloom, J.R., Fobair, P., Spiegel, D., Cox, R.S., Varghese, A. & Hoppe, R. (1991).Social supports and the social well-being of cancer survivors. *Advances in Medical Sociology*, *2*, 95-114.

Bowling, A., Farquar, M. & Grundy, E. (1996). Associations with changes in life satisfaction among three samples of elderly people living at home. *International Journal of Geriatric Psychiatry*, *11*, 1077-1087.

Bradburn, N.M. (1969). *The structure of psychological well-being*. Chicago: Aldine.

Brief, A., Butcher, A.H., George, J.M. & Link, K.E. (1993). Integrating bottom-up and top-down theories of subjective well-being: The case of health. *Personality and Social Psychology*, *64*, 646-653.

Broe, G.A., Jorm, A.F., Creasey, H., Grayson, D., Edelbrock, D., Waite, L. et al. (1998). Impact of chronic systemic and neurological disorders on disability, depression and life satisfaction. *International Journal of Geriatric Psychiatry*, *13*, 667-673.

Brown, T.A., Chorpita, B.F. & Barlow, D.H. (1998). Structural relationships among the dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology, 107*, 179-192.

Brown, T.A., Chorpita, B.F., Korotitsch, W. & Barlow, D.H. (1997). Psychometric properties of the Depression Anxiety and Stress Scales in clinical samples. *Behavior Research and Therapy, 35*, 179-192.

Browne, G., & Bramston, P. (1996). Quality of life in the families of young people with intellectual disabilities. *Australian and New Zealand Journal of Mental Health Nursing, 5*, 120-130.

Brugha, T.S., Jenkins, R., Taub, N., Meltzer, H. & Bebbington, P.E. (2001). A general population comparison of the Composite International Diagnostic Interview (CIDI) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). *Psychological Medicine, 31*, 1001-1013.

Buckley, T.C., Parker, J.D. & Heggie, J. (2001). A psychometric evaluation of the BDI-II in treatment seeking substance abusers. *Journal of Substance Abuse Treatment, 20*, 197-204.

Bushnell, J., McLeod, D., Dowell, A., Salmond, C., Ramage, S., Collings, S. et al. (2005). Do patients want to disclose psychological problems to GPs? *Family Practice*, *22*(6), 631-637.

Callahan, L.F., Kaplan, M.R. & Pincus, T. (1991). The Beck Depression Inventory, Center for Epidemiological Studies Depression Scale (CES-D), and general well-being schedule depression subscale in rheumatoid arthritis criterion contamination of responses. *Arthritis Care & Research, 4*(1), 3-11.

Camara, W.J., Nathan, J.S. & Puente, A.E. (2000). Psychological tests usage: Implications in professional psychology. *Professional Psychology, Research, 31*(2), 141-154.

Campbell, A., Converse, P.E. & Rodgers, W.L. (1976). *The quality of American life: Perceptions, evaluations, and satisfactions*. New York: Russell Safe Foundation.

Cannon, W.B. (1932). *The Wisdom of the body*. New York: WW Norton.

Cantril, H. (1965). *The pattern of human concerns*. New Jersey: Rutgers University Press.

Carver, C.S., & Schier, M.F. (1982). Control theory: A useful conceptual framework for personality-social, clinical, and health psychology. *Psychological Bulletin, 92*, 111-135.

Chaturvedi, S.K. (1991). Clinical irrelevance of HADS factor structure. *British Journal of Psychiatry*, *159*, 859-860.

Cheung, C.K. & Bagley, C. (1998). Validating an American Scale in Hong Kong: The Centre for Epidemiological Studies Depression Scale. *The Journal of Psychology*, *132*(2), 169-186.

Clara, I.P., Cox, B.J. & Enns, M.W. (2001). Confirmatory factor analysis of the Depression Anxiety and Stress Scales in depressed and anxious patients. *Journal of Psychopathology and Behavioral Assessment*, *23*, 61-67.

Coelho, R., Martins, A. & Barros, H. (2002). Clinical profiles relating gender and depressive symptoms among adolescents ascertained by the Beck Depression Inventory II. *European Psychiatry*, *17*(222-226).

Cohen, J.W. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioural sciences* (2nd ed.). New York: Erlbaum.

Colman, A.M. (2003). *Oxford Dictionary of Psychology*. New York: Oxford University Press Inc.

Costa, P.T. & McCrae, R.R. (1980). Influence of extraversion and neuroticism on subjective wellbeing: Happy and unhappy people. *Journal of Personality and Social Psychology*, *38*, 668-678.

Costa, P.T. & McCrae, R.R. (1988). Personality in adulthood: A six year longitudinal study of self-reports and spouse ratings of the NEO personality Inventory. *Journal of Personality and Social Psychology*, *54*, 853-863.

Costa, P.T. & McCrae, R.R. (1989). Personality as a lifelong determinant of well-being. In L.Z. Malatesta & C.E. Izzard (Eds.), *Emotion in adult development* (pp. 141-157). Beverley Hills, CA: Sage Publications.

Costa, P.T. & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, Florida: Psychological Assessment Resources.

Crawford, J.R. & Henry, J.D. (2003). The Depression Anxiety Stress Scale (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology, 42,* 111-131.

Crawford, J.R., Henry, J.D., Crombie, C. & Taylor, E.P. (2001). Brief report normative data for the HADS from a large non-clinical sample. *British Journal of Clinical Psychology, 40,* 429-434.

Cummins, R.A. (1995). On the trail of the gold standard for life satisfaction. *Social Indicators Research, 35,* 179-200.

Cummins, R.A. (1996). The domains of life satisfaction: An attempt to order chaos. *Soc Indicators Res, 38,* 303-328.

Cummins, R.A. (1997). *Comprehensive quality of life scale-adult manual*: *Fifth edition*. Melbourne: School of Psychology, Deakin University. (ISBN 073002804 6)

Cummins, R.A. (1998). The second approximation to an international standard of life satisfaction. *Social Indicators Research, 43,* 307-334.

Cummins, R.A. (2000a). Objective and subjective quality of life: An interactive model. *Social Indicators Research, 1,* 55-72.

Cummins, R.A. (2000b). Personal income and subjective well-being: A review. *Journal of Happiness Studies, 1,* 133-158.

Cummins, R.A. (2001). The subjective well-being of people caring for a severely disabled family member at home: A review. *Journal of Intellectual and Developmental Disability, 26,* 83-100.

Cummins, R.A. (2003). Normative life satisfaction: Measurement issues and a homeostatic model. *Social Indicators Research, 64,* 225-256.

Cummins, R.A. (2005a). *Caregivers as managers of subjective wellbeing: A homeostatic perspective*. Unpublished manuscript, Deakin University, Melbourne, Victoria, Australia.

Cummins, R.A. (2005b). *Directory of instruments*. Retrieved 10/09/05, from http://www.deakin.edu.au/research/acqol/instruments/instrument.php.

Cummins, R.A. (2009). Subjective wellbeing, homeostatically protected mood and depression. *Journal of Happiness Studies*, (in press).

Cummins, R.A., Eckersley, R., Pallant, J., Van Vugt, J. & Misajon, R. (2003). Developing a national index of subjective wellbeing: The Australian Unity Wellbeing Index. *Social Indicators Research, 64*, 159-190.

Cummins, R.A. & Gullone, E. (2000). Why we should not use 5-point Likert scales: The case for subjective quality of life measurement. In *Proceedings, Second International Conference on Quality of Life in Cities* (pp. 74-93). Singapore: National University of Singapore.

Cummins, R.A. & Lau, A.L.D. (2004). The motivation to maintain subjective well-being: A homeostatic model. In H. Switzky (Ed.), *International review of research on mental retardation: Personality and motivational systems in mental retardation* (Vol. 28, pp. 255-301). Amsterdam: Elsevier.

Cummins, R.A. & Lau, A.L.D. (2005). Quality of life measurement. In: R. Norman & D. Currow (Eds.), *Supportive care of the urology patient* (pp. 5-23). Oxford: Oxford University Press.

Cummins, R.A., Lau. A.D.L. & Davern, M. (2009). Homeostatic mechanisms and the quality of life. In K.C. Land (Ed.), *Handbook of social indicators and quality-of-life studies. Volume I: Theoretical and methodological foundations.* New York: Springer, (in press).

Cummins, R.A., McCabe, M.P., Romeo, Y. & Gullone, E. (1994). The Comprehensive Quality of Life Scale (ComQol): Instrument development and psychometric evaluation on college staff and students. *Educational and Psychological Measurement*, *54*(1), 372.

Cummins, R.A. & Nistico, H. (2002). Maintaining life satisfaction: The role of positive cognitive bias. *Journal of Happiness Studies*, *3*, 37-69.

Cummins, R.A., Gullone, E. & Lau, A.L.D. (2002). A model for subjective well-being homeostasis: The role of personality. In E. Gullone & R.A. Cummins (Eds.), *The universality of subjective well-being indicators: Social indicators research series* (pp. 7-46). Dordrecht: Kluwer.

Cummins, R.A., Walter, J. & Woerner, J. (2007). *Special report: The wellbeing of Australians: Groups with the highest and lowest wellbeing in Australia*. Melbourne: The Australian Centre on Quality of Life, School of Psychology, Deakin University. Retrieved 19/12/07, from http://www.deakin.edu.au/research/acqol/index_wellbeing/index.htm.

Cummins, R.A., Woerner, J., Gibson, A., Weinberg, M., Collard, J. & Chester, M. (2009). *Australian Unity Wellbeing Index. Report 21.0. The wellbeing of Australians: Gambling, chocolate and swine flu*. Melbourne: Australian Centre on Quality of Life, School of Psychology, Deakin University. Retrieved 06/09/09, from http://www.deakin.edu.au/research/acqol/index_wellbeing/index.htm.

Cummins, R.A., Woerner, J., Tomyn, A., Gibson, A. & Knapp, T. (2006). *Australian Unity Wellbeing Index. Report 16.0. The wellbeing of Australians: Mortgage payments and home ownership*. Melbourne: Australian Centre on Quality of Life, School of Psychology, Deakin University.

Davern, M. (2003). *Preliminary analyses for doctoral dissertation: SWB as an affective construct*. Unpublished manuscript. School of Psychology, Deakin University, Melbourne, Victoria, Australia.

Davern, M. (2004). *Subjective wellbeing as an affective construct.* Unpublished Doctoral Thesis. School of Psychology, Deakin University, Melbourne, Victoria, Australia.

Davern, M. & Cummins, R.A. (2006). Is life dissatisfaction the opposite of life satisfaction? *Australian Journal of Psychology, 58*(1), 1-7.

Davern, M., Cummins, R.A. & Stokes, M. (2007). Subjective wellbeing as an affective/cognitive construct. *Journal of Happiness Studies, 8*(4), 429-449.

Darwin, C. (1871). *The descent of man, and selection in relation to sex* (Vol. 1). London: John Murray.

Daza, P., Novy, D.M., Stanley, M.A. & Averill, P. (2002). The Depression Anxiety and Stress Scale-21: Spanish translations and validation with a Hispanic sample. *Journal of Psychopathology and Behavioral Assessment, 24*, 195-205.

De Leval, N. (1999). Quality of life and depression: Symmetry concepts. *Quality of Life Research, 8*(4), 283-291.

Depue, R.A. & Collins, P.F. (1999). Neurobiology of the structure of personality: Dopamine facilitation of incentive, motivation and extraversion. *Behavioral and Brain Sciences, 22*, 491-569.

Deinzer, R., Steyer, R., Eid, M., Notz, P., Schwenkmezger, P., Ostendorf, F. et al. (1995). Situational effects in trait assessment: The FPI, NEOFFI, and EPI questionnaires. *European Journal of Personality, 9*, 1-23.

De Neve, K.M. (1999). Happy as an extraverted clam? The role of personality for subjective well-being. *Psychological Science, 8*, 141-144.

DeNeve, K.M. & Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin, 124*, 197-229.

DeVellis, R.F. (2003). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage.

Diener (1984). Subjective Wellbeing. *Psychological Bulletin*, *95*, 542-575.

Diener, E. & Biswas-Diener, R. (2002). Will money increase subjective well-being? A literature review and guide to needed research. *Social Indicators Research*, *57*, 119-169.

Diener, E., Diener, C. & Diener, M. (1995). Factors predicting the subjective well-being of nations. *Journal of Personality and Social Psychology*, *69*, 851-864.

Diener, E., Emmons, R., Larsen, J. & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal Personality Assessment*, *49*(1), 71-75.

Diener, E. & Larsen, R.J. (1984). Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses. *Journal of Personality and Social Psychology*, *47*, 871-873.

Diener, E., Napa Scollon, C. & Lucas, R.E. (2004). The evolving concept of subjective well-being: The multifaceted nature of happiness. *Advances in Cell Aging and Gerontology*, *15*, 187-219.

Diener, E., Suh, E.M., Lucas, R.E. & Smith, H.L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, *125*, 276-303.

Dillon, K.M., Minchoff, B. & Baker, K.H. (1985). Positive emotional states and enhancement of the immune system. *International Journal of Psychiatry in Medicine*, *15*, 13-18.

Dohrenwend, B.P. (1966). Social status and psychological disorder: An issue of substance and an issue of method. *American Sociological Review*, *31*(1), 14-34.

Doyle, K.O. & Youn, S. (2000). Exploring traits of happy people. *Social Indicators Research*, *52*, 195.

Dozois, D.A. & Dobson, K.S. (2004). The prevention of anxiety and depression: Introduction. In D.J. Dozois & K.S. Dobson (Eds.), *The prevention of anxiety and depression: Theory, research, and practice* (pp. 1-6). Washington, DC: American Psychological Association.

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

Druss, B.G., Rosenheck, R.A. & Sledge, W.H. (2000). Health and disability costs of depressive illness in a major U.S. corporation. *American Journal of Psychiatry, 157,* 1274-1278.

Duffy, C., Cunningham, E.G. & Moore, S.M. (2005). The factor structure of mood states in an early adolescent sample (Brief report). *Journal of Adolescence, 28*(5), 677-680.

Dunbar, M., Ford, G., Hunt, K. & Der, G. (2000). A confirmatory factor analysis of the Hospital Anxiety and Depression Scale: Comparing empirically and theoretically derived structures. *British Journal of Clinical Psychology, 39*, 79-94.

Dunner, D.L. & Tay, L.K. (1993). Diagnostic reliability of the history of hypomania in bipolar II patients and patients with major depression. *Comprehensive Psychiatry, 34*, 303-307.

Egan, J.P. (1975). *Signal detection theory and ROC analysis, series in cognition and perception*. New York: Academic Press.

Eid, M. & Diener, E. (2004). Global judgments of subjective well-being: Situational variability and long-term stability. *Social Indictors Research, 65*, 245-277.

Einstein, D.A., Lovibond, P.F. & Gaston, J.E. (2000). Relationship between perfectionism and emotional symptoms in an adolescent sample. *Australian Journal of Psychology, 52*, 89-93.

Eysenck, H.J., Wakefield, J.A. & Friedman, A.F. (1983). Diagnosis and clinical assessment: The DSM-III. *Annual Review of Psychology*, *34*, 167–193.

Faravelli, C., Abardi, L., Bartolozzi, D., Cecchi, C., Cosci, F., D'Adamo, D. et al. (2004). The Sest Fiorentino Study: Background, methods and preliminary results. Lifetime prevalence of psychiatric disorders in an Italian community sample using clinical interviewers. *Psychotherapy and Psychosomatics*, *73*, 216-225.

Faravelli, C., Ravaldi, C. & Truglia, E. (2005). Unipolar depression. In E.J.L. Griez, C. Faravelli, D.J. Nutt & J. Zohar (Eds.), *Mood disorders: Clinical management and research issues*. Chichester, West Sussex: John Wiley & Sons.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*, 861-874.

Fayers, P.M. & Hand, D.J. (1997). Factor analysis, causal indicators and quality of life. *Quality of Life Research*, *6*(2), 139-150.

Fayers, P.M. & Hand, D.J. (2002). Causal variables, indicator variables and measurement scales: An example from quality of life. *Journal of the Royal Statistical Society, 165*(2), 233-261.

Ferring, D., Filipp, S.H. & Schmidt, K. (1996). The Skala zur Lebensbewertung: Scale construction and findings on reliability, stability and validity. *Zeitschrift fur Differentielle und Diagnostiche Psychologie*, *17*, 141-153.

First, M.B., Spitzer, R.L. & Williams, J.B.W. (1990). Exclusionary principles and the comorbidity of psychiatric diagnosis: A historical review and implications for the future. In J.D. Maser & C.R. Cloninger (Eds.), *Comorbidity of mood and anxiety disorders* (Vol. 83-109). Washington, DC: American Psychiatric Press.

Flett, G.L., Vrendenburg, K. & Krames, L. (1997). The continuity of depression in clinical and nonclinical samples. *Psychological Bulletin*, *121*, 395–416.

Fluss, R., Faraggi, D. & Reiser, B. (2005). Estimation of the Youden Index and its associated cut off point. *Biometrical Journal*, *47*, 458-472.

Frances (1998). Problems in defining clinical significance in epidemiological studies. *Archives of General Psychiatry*, *55*, 119.

Franklin, L.C., Strong, D.R. & Greene, R.L. (2002). A taxonometric analysis of the MMPI-2 depression scales. *Journal of Personality Assessment*, *79*(1), 110-121.

Fredrickson, B.L. & Joiner, T. (2002). Positive emotions trigger upward spirals toward emotional well-being. *Psychological Science*, *13*, 172-175.

Gloster, A.T., Rhoades, H.M., Novy, D.M., Klotsche, J., Senior, A., Kunik, M. et al. (1998). Psychometric properties of the Depression Anxiety and Stress Scale-21 in older primary care patients. *Journal of Affective Disorders*, *110*, 248-259.

Goldberg, J. & Harrow, M. (2005). Subjective life satisfaction and objective functional outcome in bipolar and unipolar mood disorders: A longitudinal analysis. *Journal of Affective Disorders*, *89*(1), 79-89.

Golberg, D.P. & Lecrubier, Y. (1995). Form and frequency of mental disorders across Centres. In T.B. Üstün & N. Sartorius (Eds.), *Mental illness in general health care: An international study* (pp. 323-334). Chichester: John Wiley & Sons.

Goodwin, D.W. & Guze, S.B. (1989). *Psychiatric diagnosis* (4th ed.). New York: Oxford University Press.

Goldings, H.J. (1954). On avowal and projection of happiness. *Journal of Personality*, *23*, 30-47.

Greenberg, P.E., Kessler, R.C., Birnbaum, H.G., Leong, S.A., Lowe, S.W., Berglund, P.A. et al. (2003). The economic burden of depression in the United States: How did it change between 1990 and 2000? *Journal of Clinical Psychiatry*, *64*(12), 1465-1475.

Greenberg, P.E., Stiglin, L.E., Finkelstein, S.N. & Berndt, E.R. (1993). Depression: A neglected major illness. *Journal of Clinical Psychiatry*, *54*, 419-424.

Groot, W. & VandenBrink, M. (2000). Life satisfaction and preference drift. *Social Indicators Research*, *50*, 315-329.

Grothe, K.B., Dutton, G.R., Jones, G.N., Bodenlos, J., Ancona, M. & Brantley, P.J. (2005). Validation of the Beck Depression Inventory-II in a low-income African American sample of medical outpatients. *Psychological Assessment*, *17*, 110-114.

Groth-Marnat, G. (2003). *Handbook of psychological assessment* (4th ed.). New York: John Wiley & Sons, Incorporated.

Gunderson, J.G., Links, P.S. & Reich, J.H. (1991). Competing models of personality disorders. *Journal of Personality Disorders*, *5*, 60-68.

Guze, S.B. (1992). *Why psychiatry is a branch of medicine*. New York: Oxford University Press.

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

Hanested, B.R. & Albrektsen, G. (1992). The stability of quality of life experience in people with Type 1 diabetes over a period of a year. *Journal of Advanced Nursing*, *17*, 777-784.

Hanley, J.A. & McNeil, M.D. (1982). The meaning and use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology*, *143*, 29-36.

Hansson, L. (2002). Quality of life and depression. *International Review of Psychiatry*, *14*, 185-189.

Harman, H.H. (1976). *Modern factor analysis* (3rd ed.). Chicago: University of Chicago Press.

Hawthorne, G., Cheok, F., Goldney, R. & Fisher, L. (2003). The excess cost of depression in South Australia: A population-based study. *Australian and New Zealand Journal of Psychiatry, 37*, 362-373.

Headey, B.W., Kelley, J. & Wearing, A.J. (1993). Dimensions of mental health: Life satisfaction, positive affect, anxiety and depression. *Social Indicators Research, 29*(1), 63-82.

Headey, B. & Wearing, A. (1988). The sense of relative superiority – central to well-being. *Social Indicators Research, 20,* 497-516.

Headey, B. & Wearing, A. (1989). Personality, life events, and subjective well-being: Toward a dynamic equilibrium model. *Journal of Personality and Social Psychology, 57,* 731-739.

Headey, B. & Wearing, A. (1992). *Understanding happiness: A theory of subjective well-being.* Melbourne, VIC: Longman Cheshire.

Henry, J.D. & Crawford, J.R. (2005). The short-form version of the Depression Anxiety and Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 44,* 227-239.

Hermann, C. (1997). International experiences with the Hospital Anxiety and Depression Scale – A review of validation data and clinical results. *Journal of Psychosomatic Research, 42,* 17-41.

Hofstede, G. (1980). *Cultures consequences: International differences in work-related values*. Newsbury Park, CA: Sage.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6,* 65-70.

Hong, S.M. & Giannikopoulos, E. (1994). The relationship of satisfaction with life to personality characteristics. *The Journal of Psychology, 128*(5), 547-559.

Hooke, G.R. & Page, A.C. (2002). Predicting outcomes of group cognitive behavior therapy for patients with affective and neurotic disorders. *Behavior Modification*, *26*(5), 648-658.

Ingram, R.E. & Seigle, G.J. (2002). Contemporary methodological issues in the study of depression. In I.H. Gotlib & C.L. Hammen (Eds.), *Handbook of depression* (pp. 86-114). New York: Guildford Press.

International Wellbeing Group. (2006). *Personal Wellbeing Index Manual*. Melbourne: Deakin University. Retrieved 30/06/09, from http://www.deakin.edu.au/research/acqol/instruments/wellbeing_index.htm.

Iwamasa, G.Y., Hilliard, K.M. & Osato, S.S. (1998). Conceptualising anxiety and depression: The Japanese American older adult perspective. *Clinical Gerontologist*, *19*, 77-93.

Jacobi, F., Wittchen, H.U., Holting, C., Hofler, M., Muller, N., Pfister, H. et al. (2004). Prevalence, comorbidity and correlates of mental disorders in the general population: Results from the German Health Interview and Examination Survey (GHS). *Psychological Medicine, 34*, 597-611.

Jahoda, M. (1958). *Current conceptions of positive mental health*. New York: Basic Books.

Jefferson, T. (1776). *The Declaration of Independence*. United States of America.

Jho, M.Y. (2001). Study on the correlation between depression and quality of life for Korean women. *Nursing & Health Sciences, 3*(3), 131-137.

Jones, L.V. & Thurstone, L.L. (1955). The psychophysics of semantics: An experimental investigation. *The Journal of Applied Psychology*, *39*(1), 31-36.

Judd, L.L., Akiksal, H.S. & Paulus, M.P. (1997). The role and clinical significance of Subsyndromal Depressive Symptoms (SSD) in unipolar major depressive disorder. *Journal of Affective Disorders, 45*, 5-18.

Judd, L.L., Rappaport, M.H., Paulus, M.P. & Brown, J L. (1994). Subsyndromal symptomatic depression: A new mood disorder? *Journal of Clinical Psychiatry*, *5*(4), 18-28.

Kaiser, H. (1970). A second generation Little Jiffy. *Psychometrika, 35*, 401-415.

Kaiser, H. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31-36.

Kasser, T. & Ryan, R.M. (1996). Further examining the American dream: Differential correlates of intrinsic and extrinsic goals. *Personality and Social Psychology Bulletin, 22,* 280-287.

Kendall, P.C. & Watson, D. (1989). *Anxiety and depression: Distinctive and overlapping features*. San Diego, CA: Academic Press.

Keogh, E. & Cochrane, M. (2002). Anxiety sensitivity, cognitive biases, and the experience of pain. *Journal of Pain*, *3*(4), 320-329.

Kessler, R.C., Abelson, J., Demler, O., Escobar, J.I., Gibbon, M., Guyer, M.E. et al. (2004). Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMH-CIDI). *International Journal of Methods in Psychiatric Research, 13*, 122-139.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Koretz, D. & Merikangas, K.R. (2003). The epidemiology of Major Depressive Disorder. *Journal of the American Medical Association, 289*, 3095-3105.

Kessler, R.C., Gotlib, I.H. & Hammen, C.L. (2002). Epidemiology of depression. In I.H. Gotlib & C.L. Hammen (Eds.), *Handbook of depression* (pp. 23-42). New York, NY: Guilford Press.

Kessler, R.C. & Üstün, T.B. (2004). The World Mental Health (WMH) survey initiative version of the World Health Oraganization's Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research, 13*(2), 93-121.

King, K.M. (2000). Gender and short-term recovery from cardiac surgery. *Nursing Research, 49*, 29-36.

Koivumaa-Honkanen, H., Honkanen, R., Viinamaeki, H., Heikkilae, K., Kaprio, J. & Koskenvuo, M. (2001). Life satisfaction and suicide: A 20-year follow-up study. *American Journal of Psychiatry, 158*(3), 433-439.

Koivuuma-Honkanen, H., Kaprio, J., Honkanen, R., Viinamäki, H. & Koskenvuo, M. (2004). Life satisfaction and depression in a 15-year follow-up of healthy adults. *Social Psychiatry and Psychiatric Epidemiology, 39*(12), 994-999.

Komiti, A.A., Jackson, H.J., Judd, F.K., Cockram, A.M., Kyrios, M., Yeatman, R. et al. (2001). A comparison of the Composite International Diagnostic Interview (CIDI-auto) with clinical assessment in diagnosing mood and anxiety disorders. *Australian and New Zealand Journal of Psychiatry, 35,* 224-230.

Kozma, A., Stone, S. & Stones, M. (2000). Stability in components and predictors of Subjective Well-being (SWB): Implications for SWB structure. In E. Diener (Ed.), *Advances in quality of life theory and research*. Boston: Kluwer Academic Publishers.

Kringlen, E., Torgersen, S. & Cramer, V. (2001). A Norwegian psychiatric epidemiological study. *The American Journal of Psychiatry, 158,* 1091-1098.

Kumar, G., Steer, R.A., Teitelman, K.B. & Villacis, L. (2002). Effectiveness of Beck Depression Inventory-II subscales in screening for major depressive disorders in adolescent psychiatric inpatients. *Assessment, 9*(2), 164-170.

Kurdyak, P.A. & Gnam, W.H. (2005). Small signal, big noise: Performance of the CIDI depression module. *Canadian Journal Of Psychiatry. Revue Canadienne De Psychiatrie, 50*(13), 851-856.

Larsen, R. J. (2000). Toward a science of mood regulation. *Psychological Inquiry, 11*(3), 129-141.

Larsen, R.J. & Diener, E. (1987). Affect intensity as an individual difference characteristic: A review. *Journal of Research Personality*, *21*, 1-39.

Larsen, R.J., Diener, E. & Emmons, R.A. (1985). An evaluation of subjective well-being measures. *Social Indicators Research*, *17*, 1-17.

Lecrubier, Y. & Üstün, T.B. (1998). Panic and depression: A worldwide primary care perspective. *International Clinical Psychopharmacology*, *13*(4), S7-SS11.

Lewinsohn, P.M., Hoberman, H.M. & Rosenbaum, M. (1988). A prospective study of risk factors for unipolar depression. *Journal of Abnormal Psychology*, *97*(3), 251-264.

Lewinsohn, P.M., Redner, J.E. & Seeley, J.R. (1991). The relationship between life satisfaction and psychosocial variables: New perspectives. In F. Strack, M. Argyle & N. Schwartz (Eds.), *Subjective well-being: An interdisciplinary perspective*. UK: Pergamon Press.

Lewis, C.A., Dorahy, M.J. & Schumaker, J.F. (1999). Depression and life satisfaction among Northern Irish adults. *Journal of Social Psychology*, *139*(4), 533-535.

Lima, M.S., Beria, J.U., Tomasi, E., Conceicao, A.T. & Mari, J.J. (1996). Stressful life events and minor psychiatric disorders: An estimate of the population attributable fraction in a Brazilian community based study. *International Journal of Psychiatry and Medicine*, *26*, 211-222.

Lindquist, R., Dupuis, G., Hoogwerf, B., Czajkowski, S., Herd, A.J., Barton, F. et al. (2003). Comparison of health-related quality-of-life outcomes for men and women after coronary artery bypass surgery through 1 year: Findings from the POST CABG Biobehavioural Study. *American Heart Journal*, *146*, 1038-1044.

Lisspers, J., Nygren, A. & Soderman, E. (1997). Hospital Anxiety and Depression Scale (HAD): Some psychometric data for a Swedish sample. *Acta Psychiatrica Scandinavia*, *96*, 281-286.

Littlewood, R. (1990). A decade of the "new cross-cultural" psychiatry. *British Journal of Psychiatry*, *156*, 308-327.

Lönquvist, J. (2000). Psychiatric aspects of suicidal behavior: Depression. In Hawton, K., Van Herrington, K. (Eds.), *The international handbook of suicide and attempted suicide* (pp. 107-120). Chichester, UK: John Wiley.

Lovibond, P.F. & Lovibond, S.H. (1995b). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335-343.

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.

Lusted, L.B. (1971). Decision-making studies in patient management. *New England Journal of Medicine*, *284*, 416-424.

Lyubomirsky, S., Sheldon, K.M. & Schkade, D. (2005). Pursuing happiness: The architecture of sustainable change. *Review of General Psychology, 9*(2), 111-131.

Maier, W., Ganisicke & Weiffenbach, O. (1997). The relationship between major and subthreshold variants of unipolar depression. *Journal of Affective Disorders*, *45*, 41-51.

Marks, G.N. & Fleming, N. (1999). Influences and consequences of well-being among Australian young people: 1980–1995. *Social Indicators Research*, *46*, 301-323.

Martin, C.R. & Newell, R.J. (2004). Factor structure of the Hospital Anxiety and Depression Scale in individuals with facial disfigurement. *Psychology, Health, and Medicine*, *9*, 327-336.

Martin, C.R., Tweed, A.E. & Metcalfe, M.S. (2004). A psychometric evaluation of the Hospital Anxiety and Depression Scale in patients diagnosed with end-stage renal disease. *British Journal of Clinical Psychology, 43*, 51-64.

McKendrick-MacKay, Y. (2008). *The Processes through which affect infuses judgements of subjective wellbeing.* Unpublished Doctoral Thesis. Deakin University, Melbourne, Victoria, Australia.

Meakin, C.J. (1992). Screening for depression in the medically ill: The future of paper and pencil tests. *British Journal of Psychiatry, 160,* 212-216.

Metz, C.E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine, 8*, 283-298.

Mezzich, J.E., Kirmayer, L.J., Kleinmann, A., Fabrega, H.J., Parron, D.L., Good, B.J. et al. (1999). The place of culture in DSM-IV. *The Journal of Nervous and Mental Disease, 187*(8), 457-464.

Michalos, A.C. (1985). Multiple Discrepancies Theory (MDT). *Social Indicators Research, 16*, 349-373.

Mirowsky, J. (1994). The advantages of indexes over diagnoses in scientific assessment. In W.R. Avison & I.H. Gotlib (Eds.), *Stress and mental health: Contemporary issues and prospects for the future* (pp. 239–258). New York: Plenum.

Moorey, S., Greer, S., Watson, M., Gorman, C., Rowden, L., Tunmore, R. et al. (1991). The factor structure and factor stability of the Hospital Anxiety and Depression Scale in patients with cancer. *British Journal of Psychiatry, 58*, 255-259.

Mulrow, C.D., Williams, J.W., Gerrety, M.B., Ramirez, G., Montiel, O.M. & Kerber, C. (1995). Case-finding instruments for depression in primary care settings. *Annals of Internal Medicine, 122,* 913-921.

Murray, C.J.L. & Lopez, A.D. (1996). *The global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected until 2020* (Vol. 1). Cambridge, MA: Harvard School of Public Health on behalf of the World Health Organisation and the World Bank.

Murray, C.J.L. & Lopez, A.D. (2000). Progress and directions in refining the global burden of disease approach: A response to Williams. *Health Economics, 9*, 69-82.

Murray, C.J.L., Lopez, A.D., Mathers, C.D. & Stein, C. (2001). *The global burden of disease 2000 project: Aims, methods and data sources*. Global Programme on Evidence for Health Policy Discussion Paper No. 36. Retrieved from http://www.who.int/healthinfo/paper36.pdf.

Narrow, W.E., Rae, D.S., Robins, L.N. & Regier, D.A. (2002). Revised prevalence estimates of mental in the United States. Using a clinical significance criterion to reconcile two surveys' estimates. *Archives of General Psychiatry*, *59*, 115-123.

Nezu, A.M., Ronan, G.F., Meadows, E.A. & McClure, K.S. (2000). *Clinical assessment series: Vol. 1. Practitioner's guide to empirically based measures of depression*. New York, NY: Kluwer/Plenum.

Nieuwenhuijsen, K., de Boer, A.G., Verbeek, J.H., Blonk, R.W. & Van Dijk, F.J. (2003). The Depression Anxiety and Stress Scales (DASS): Detecting anxiety disorder and depression in employees absent from work because of mental health problems. *Occupational and Environmental Medicine, 60*, 177-182.

Norton, P.J. (2007). Depression Anxiety and Stress Scales (DASS-21): Psychometric analysis across four racial groups. *Anxiety, Stress, and Coping*, *20*(3), 253-265.

Norton, P.J., Hayes, S.A. & Hope, D.A. (2004). Effects of a transdiagnostic group treatment for anxiety and secondary depressive disorders. *Depression and Anxiety*, *20*, 198-202.

Norton, P.J. & Hope, D.A. (2005). Preliminary evaluation of a broad-spectrum cognitive behavioral group therapy for anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *36*, 79-97.

Okun, M.A., Stock, W.A., Haring, M.J. & Witter, R.A. (1984). The social activity/subjective well-being relation: A quantitative synthesis. *Research on Aging*, *6*, 45-65.

Ostir, G.V., Markides, K.S., Black, S.A. & Goodwin, J.S. (2000). Emotional well-being predicts subsequent functional independence and survival. *Journal of the American Geriatric Society*, *48*, 473-478.

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

Osman, A., Kopper, B.A., Barrios, F., Gutierrez, P.M. & Bagge, C.L. (2004). Reliability and validity of the Beck Depression Inventory-II with adolescent psychiatric inpatients. *Psychological Assessment*, *16*(2), 120-132.

Pallant, J. (2007). *SPSS survival manual* (3rd ed.).Sydney: Allen & Unwin.

Pallant, J.F. & Tennant, A. (2007). An introduction to the Rasch measurment model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, *46*, 1-18.

Patterson, D., Everett, J.J., Bombardier, C.H., Questad, K.A., Lee, V.K. & Marvin, J.A. (1993). Psychological effects of severe burn injuries. *Psychological Bulletin*, *113*, 362-378.

Pavot, W. & Diener, E. (1993). Review of the satisfaction with life scale. *Psychological Assessment*, *5*, 164-172.

Peat, J. (2001). *Health science research: A handbook of quantitative methods*. Sydney, NSW: Allen & Unwin.

Persons, J.B. (1986). The advantages of studying psychological phenomena rather than psychiatric diagnoses. *American Psychologist*, *41*, 1252-1260.

Pincus, H.A., Davis, W.W. & McQueen, L.E. (1999). 'Subthreshold' mental disorders. A review and synthesis of studies on minor depression and other 'brand names'. *The British Journal of Psychiatry*, *174*, 288-296.

Regier, D.A., Kaelber, C.T., Rae, D.S., Farmer, M.E., Knauper, B., Kessler, R.C. et al. (1998). Limitations of diagnostic criteria and assessment instruments for mental disorders. Implications for research and policy. *Archives of General Psychiatry*, *55*, 109-115.

Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161-1178.

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

Ritterband, L.M. & Spielberger, C.D. (1996). Construct validity of the Beck Depression Inventory as a measure of state and trait depression in nonclinical populations. *Depression and Stress*, *2*, 123-145

Ritterband, L.M. & Spielberger, C.D. (2001). Depression in a cancer patient population. *Journal of Clinical Psychology in Medical Settings*, *8*(2), 85-93.

Robins, L.N. & Helzer, J.E. (1986). Diagnosis and clinical assessment: The current state of psychiatric diagnosis. *Annual Review of Psychology*, *37*, 409–432.

Rogers, W., Adler, D., Bungay, K. & Wilson, I. (2005). Depression screening instruments made good severity measures in a cross-sectional analysis. *Journal of Clinical Epidemiology*, *58*, 370-377.

Ruopp, M.D., Perkins, N.J., Whitcomb, B.W. & Schisterman, E.F. (2008). Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biometrical Journal*, *50*(3), 419-430.

Ruscio, J. & Ruscio, A.M. (2000). Informing the continuity controversy: A taxometric analysis of depression. *Journal of Abnormal Psychology, 109*(3), 473-487.

Ruscio, J. & Ruscio, A.M. (2002). A structure-based approach to psychological assessment. *Assessment*, *9*(1), 4-16.

Rush, A.J., Gullion, C.M., Basco, M.R., Jarrett, R.B. & Trivedi, M.H. (1996). The Inventory of Depressive Symptomatology (IDS): Psychometric properties. *Psychological Medicine*, *26*, 477-486.

Russell, J.A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145-172.

Russell, J.A. (2009). Emotion, core affect, and psychological construction. *Cognition & Emotion*, 1464-1600.

Russell, J. & Carroll, J. (1999a). On the bipolarity of positive and negative affect. *Psychological Bulletin*, *125*, 3-30.

Russo, J., Roy-Byrne, P., Reeder, D., Alexander, M., Dwyer-O'Connor, E.D., Dagadakis, C. et al. (1997). Longitudinal assessment of quality of life in acute psychiatric inpatients: Reliability and validity. *Journal of Nervous and Mental Disease*, *185*(3), 166-175.

Rusting, C.L. & Larsen, R.J. (1997). Extraversion, neuroticism, and susceptibility to positive and negative affect: A test of two theoretical models. *Personality and Individual Differences*, *22*(5), 607-612.

Sartorius, N., Üstün T.B., Lecrubier, Y. & Wittchen, H.U. (1996). Depression comorbid with anxiety: Results from the WHO study on 'Psychological disorders in primary health care.' *British Journal of Psychiatry*, *168*(30), 38-43.

Schimmack, U., Shigehiro, O., & Diener, E. (2002). Cultural influences on the relation between pleasant emotions and unpleasant emotions: Asian dialectic philosophies or individualism-collectivism. *Cognition and Emotion*, *16*(6), 705-719.

Schulenberg, S.E. & Yutrzenka, B.A. (2001). Equivalence of computerised and conventional versions of the Beck Depression Inventory-II (BDI-II). *Current Psychology*, *20*, 216-230.

Schwarz, N. & Strack, F. (1991). Evaluating one's life: A judgment model of subjective well-being. In F. Strack, M. Argyle & N. Schwarz (Eds.), *Subjective well-being* (pp. 27-47). Oxford: Pergamon.

Schwarz, N. & Strack, F. (1999). Reports of subjective well-being: Judgmental processes and their methodological implications. In D. Kahneman, E. Diener & N. Schwarz (Eds.), *Foundations of hedonic psychology: Scientific perspectives on enjoyment and suffering* (pp. 61-84). New York, NY: Russell Sage Foundation.

Segal, D.L., Hersen, M. & Van Hasselt, V.B. (1994). Reliability for the structured clinical interview for DSM-II-R: An evaluative review. *Comprehensive Psychiatry*, *35*, 316-325.

Shade, C.P., Jones, E.R.J. & Wittlin, B.J. (1998). A ten-year review of the validity and clinical utility of depression screening. *Psychiatric Services*, *49*, 55-61.

Shankman, S.A. & Klein, D.N. (2002). Dimensional diagnosis of depression: Adding the dimension of course to severity, and comparison to the DSM. *Comprehensive Psychiatry*, *43*(6), 420-426.

Shea, T.L., Tennant, A. & Pallant, J.F. (2009). Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS). *BMC Psychiatry*. Retrieved from http://www.biomedcentral.com/1471-244X/9/21. doi:10.11861/1471-244X-9-21.

Sheldon, K.M. & Houser-Marko, L. (2001). Self-concordance, goal-attainment, and the pursuit of happiness: Can there be an upward spiral? *Journal of Personality and Social Psychology*, *80*, 152-165.

Silva, P. (2005). Intellectuals, technocrats and social change in Chile: Past, present and future perspectives. In A. Angell & B. Pollack (Eds.), *The legacy of dictatorship: Political, economic and social change in Pinochet's Chile* (pp. 198-223). Liverpool: Institute of Latin American Studies.

Simon, G.E., Goldberg, D.P., Von Korff, M. & Üstün, T.B. (2002). Understanding cross-national differences in depression prevalence. *Psychological Medicine*, *32*, 585-594.

Simpson, P.L., Schumaker, J.F., Dorahy, M.J. & Shrestha, S.N. (1996). Depression and life satisfaction in Nepal and Australia. *Journal of Social Psychology*, *136*(6), 783-790.

Solomon, A., Haaga, D.A.F. & Arnow, B.A. (2001). Is clinical depression distinct from subthreshold depressive symptoms? A review of the continuity issue in depression research. *The Journal of Nervous and Mental Disease*, *189*(8), 498-506.

Spangenberg, J.J. & Theron, J.C. (1999). Stress and coping strategies in spouses of depressed patients. *Journal of Psychology*, *133*, 253-262.

Spinhoven, P. & Van der Does, A. J. (1997). Somatization and somatosensory amplification in psychiatric outpatients: An explorative study. *Comprehensive Psychiatry*, *38*, 93-97.

Sprinkle, S.D., Lurie, D., Insko, S.L., Atkinson, G., Jones, G.L., Logan, A.R. et al. (2002). Criterion validity, severity cut scores, and test-retest reliability of the Beck Depression Inventory-II in a University counselling centre sample. *Journal of Counseling Psychology, 49*(3), 381-385.

Spitzer, R.L., Kroenke, K. & Williams, J.B.W. (1999). Validation and utility of a self-report version of the PRIME-MD: The PHQ Primary Care Study. *The Journal of the American Medical Association, 282*, 1737-1744.

Spitzer, R.L., Williams, J.B.W., Gibbon, M. & First, M.B. (1990). *Structured Clinical Interview for DSM-III-R*. Washington, DC: American Psychiatric Press.

Spitzer, R.L., Williams, J.B., Gibbon, M. & First, M.B. (1992). The structured clinical interview for DSM-III-R (SCID). 1: History, rationale and description. *Archives of General Psychiatry, 49*, 624-629.

Staw, B.M., Sutton, R.I. & Pelled, L.H. (1995). Employee positive emotion and favorable outcomes at the workplace. *Organization Science, 5*, 51-71.

Steer, R.A., Beck, A.T. & Garrison, B. (1986). Applications of the Beck Depression Inventory. In N. Sartorius & T.A. Ban (Eds.), *Assessment of depression* (pp. 121-142), New York: Springer-Verlag.

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

Steer, R.A., Kumar, G., Ranieri, W.F. & Beck, A.T. (1998). Use of the Beck Depression Inventory-II with adolescent psychiatric outpatients. *Journal of Psychopathology and Behavioral Assessment, 20*(2), 127-137.

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

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

Steyer, R., Ferring, D. & Schmitt, M. (1992). On the definition of states and traits. *European Journal of Psychological Assessment, 2*, 79-98.

Stokes, M. (2007). *Calculator – slope comparison*. Unpublished Statistical calculator. School of Psychology, Deakin University, Melbourne, Victoria, Australia.

Stuckey, J.C., Neundorfer, M.M. & Smyth, K.A. (1996). Burden and well-being: The same coin or related currency? *Gerontologist, 36*, 686-693.

Suh, E., Diener, E. & Fujita, F. (1996). Events and subjective well-being: Only recent events matter. *Journal of Personality and Social Psychology, 70*, 1091-1102.

Swets, J.A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*. Hillsdale, NJ: Lawrence Erlbaum Associates, Incorporated.

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

Tatarkiewicz, W. (1976). *Analysis of happiness*. The Hague: Martinus Nijhoff.

Taylor, S.E. Brown., J.D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*, 193-210.

Tully, P.J., Zajac, I.T. & Vennings, A.J. (2009). The structure of anxiety and depression in a normative sample of younger and older Australian adolescents. *Journal of Abnormal Child Psychology*, *37*(5), 717-726.

Tymchuk, A.J. (1994). Depression symptomatology in mothers with mild intellectual disability: An exploratory study. *Australian and New Zealand Journal of Developmental Disabilities, 19*, 111-119.

Unger, J.P., De Paepe, P., Cantuarias, G.S. & Herrera, O.A. (2008). Chile's neoliberal health reform: An assessment and a critique. *PLOS Medicine, 5*(4), 542-546. Retrieved from http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0050079.

Upadhaya, A.K. & Stanley, I. (1993). Hospital anxiety depression scale: A current perspective. *British Journal of General Practice, 43*, 349-350.

Üstün, T.B. & Sartorius, N. (1995). *Mental Illness in General Health Care across the world. An international study.* New York, NY: John Wiley & Sons.

Vitterso, J. (2001). Personality traits and subjective well-being: Emotional stability, not extraversion, is probably the important predictor. *Personality and Individual Differences, 31*, 903-914.

Vitterso, J. & Nilsen, F. (2002). The conceptual and relational structure of subjective well-being, neuroticism, and extraversion: Once again, neuroticism is the important predictor of happiness. *Social Indicators Research, 57*, 89-118.

Vrendenberg, K., Flett, G.L. & Krames, L. (1993). Analogue versus clinical depression: A critical reappraisal. *Psychological Bulletin, 113*, 327-344.

Watson, D. & Clark, L.A. (1992). On traits and temperament: General and specific factors of emotional experience and their relation to the five-factor model. *Journal of Personality, 60*, 441-475.

Watson, D. & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*(2), 219-235.

Watson, D. & Walker, L.M. (1996). The long-term stability and predictive validity of trait measures of affect. *Journal of Personality and Social Psychology*, *70*, 567-577.

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.

Weary, G., Edwards, J.A. & Jacobson, J.A. (1995). Depression research methodologies in the Journal of Personality and Social Psychology: A reply. *Journal of Personality and Social Psychology*, *68*, 885-891.

Weismann, M.M., Bland, R.C., Canino, G.J., Faravelli, C., Greenwald, S., Hwu, H.G. et al. (1996). Cross-national epidemiology of major depression and bipolar disorder. *Journal of the American Medical Association*, *276*, 292-299.

Weismann, M.M., Bruce, L.M., Leaf, P.J., Florio, L.P. & Holzer III, C. (1991). Affective disorders. In L.N. Robins & D.A. Regier (Eds.), *Psychiatric disorders in America: The epidemiological catchment area study* (pp. 53-80). New York, NY: Free Press.

Whisman, M.A., Perez, J.E. & Ramel, W. (2000). Factor structure of the Beck Depression Inventory-Second Edition (BDI-II) in a student sample. *Journal of Clinical Psychology*, *56*, 545-551.

Whitlatch, C.J., Feinberg, L.F. & Sebasta, D.S. (1997). Depression and health in family caregivers. *Journal of Aging and Health*, *9*, 222-243.

Widiger, T.A. (1997). Mental disorders as discrete clinical conditions: Dimensional versus categorical classification. In S.M. Turner & M. Hersen (Eds.), *Adult psychopathology and diagnosis* (pp. 3–23). New York, NY: Wiley.

Williams, J.B., Pignone, M., Ramirez, G. & Stellato, C. (2002). Identifying depression in primary care: A literature synthesis of case-finding instruments *General Hospital Psychiatry*, *24*, 225-237.

Wittchen, H.U. (1994). Reliability and validity studies of the WHO – Composite International Diagnostic Interview (CIDI): A critical review. *Journal of Psychiatric Research*, *28*(1), 57-84.

Wittchen, H.U., Nelson, C.B. & Lachner, G. (1998). Prevalence of mental disorders and psychosocial impairments in adolescents and young adults. *Psychological Medicine, 28,* 109-126.

Williams, J.B.W., Gibbon, M., First, M.B., Spitzer, R.L., Davies, M., Borus, J. et al. (1992). The structured clinical interview for DSM-III-R (SCID). Multisite test-retest reliability. *Archives of General Psychiatry, 20,* 1-11.

Williams, J.W., Kerber, C.A., Mulrow, C.D., Medina, A. & Aguilar, C. (1995). Depressive disorders in primary care: Prevalence, functional disability, and identification. *Journal of General Internal Medicine, 10,* 7-12.

World Health Organization. (1946). *Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference*. New York, (Official Records of the World Health Organization, no. 2, p. 100)

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines.* Geneva: World Health Organization.

World Health Organization. (1997). *Composite international diagnostic interview, version 2.1*. Geneva: World Health Organization

World Health Organization. (2001). *The World Health Report 2001. Mental health: New understanding, new hope.* Geneva: World Health Organisation. Retrieved 06/01/08, from http://www.who.int/whr/2001/en/.

World Health Organization. (2005). *Depression*. Retrieved 03/9/05, from http://www.who.int/mental_health/management/depression/definition/en/print.html.

Youden, W.J. (1950). Index for rating diagnostic tests, *Cancer, 3,* 32-35.

Zigmond, A.S. & Snaith, R. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica, 67,* 361-370.

Zar, J.H. (1999). *Biostatistical analysis* (4th ed.). London, UK: Prentice Hall International.

Zimmerman, M. & Coryell, W. (1987). The Inventory to Diagnose Depression (IDD): A self-report scale to diagnose major depressive disorder. *Journal of Consulting and Clinical Psychology*, *55*, 55-59.

Zung, W.W.K. (1965). A self-rating depression scale. *Archives of General Psychiatry*, *12*, 63-70.

Zweig, M.H. & Campbell, G.C. (1993). Receiver-operating Characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, *39*, 561-577.

Appendix 1: DSM-IV-TR Criteria for Major Depressive Episode

A. Five (or more) of the following symptoms have been present during the same two-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

Note: Do not include symptoms that are clearly due to a general medical condition or mood-incongruent delusions or hallucinations.

1. depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful) **Note:** in children and adolescents, can be irritable mood
2. markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation of others)
3. significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase of appetite nearly every day **Note:** in children, consider failure to make expected weight gains
4. insomnia or hypersomnia nearly every day
5. psychomotor agitation or retardation nearly every day (observable by others not merely subjective feelings of restlessness or being slowed down)
6. fatigue or loss of energy nearly every day
7. feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick)
8. diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or observation by others)
9. recurrent thought of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide

B. The symptoms do not meet criteria for a Mixed Episode.

C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

D. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g. hypothyroidism).

E. The symptoms are not better accounted for by Bereavement, i.e., after the loss of a loved one, the symptoms persist for longer than 2 months or are characterised by marked functional impairment, morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms, or psychomotor retardation.

American Psychiatric Association (2000) *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition, Text Revision, Washington, DC: American Psychiatric Association.

Appendix 2: Ethics Approval

Research Services
Office of the Deputy Vice-Chancellor (Research) (Melbourne Campus)

MEMORANDUM

TO: A/Prof. David Mellor
School of Psychology, Burwood
cc: Renee Bear

FROM: Secretary, Deakin University Human Research Ethics Committee (DU-HREC)

DATE: 6 March 2008

SUBJECT: **Project EC 25-2008** (*Please quote this project number in future communication.*)
Subjective wellbeing and depression in Australia: A longitudinal study involving people in remote locations

---

This application was considered at the DU-HREC meeting held on 18 February 2008.

Approval has been given for Renee Bear, under the supervision of A/Prof. David Mellor, School of Psychology, to undertake this project for a period of three years from 6 March 2008.

The approval given by the Deakin University Human Research Ethics Committee is given only for the project and for the period as stated in the approval. It is your responsibility to contact the Executive Officer immediately should any of the following occur:

- Serious or unexpected adverse effects on the participants
- Any proposed changes in the protocol, including extensions of time.
- Any events which might affect the continuing ethical acceptability of the project.
- The project is discontinued before the expected date of completion.
- Modifications are requested by other HREC's.

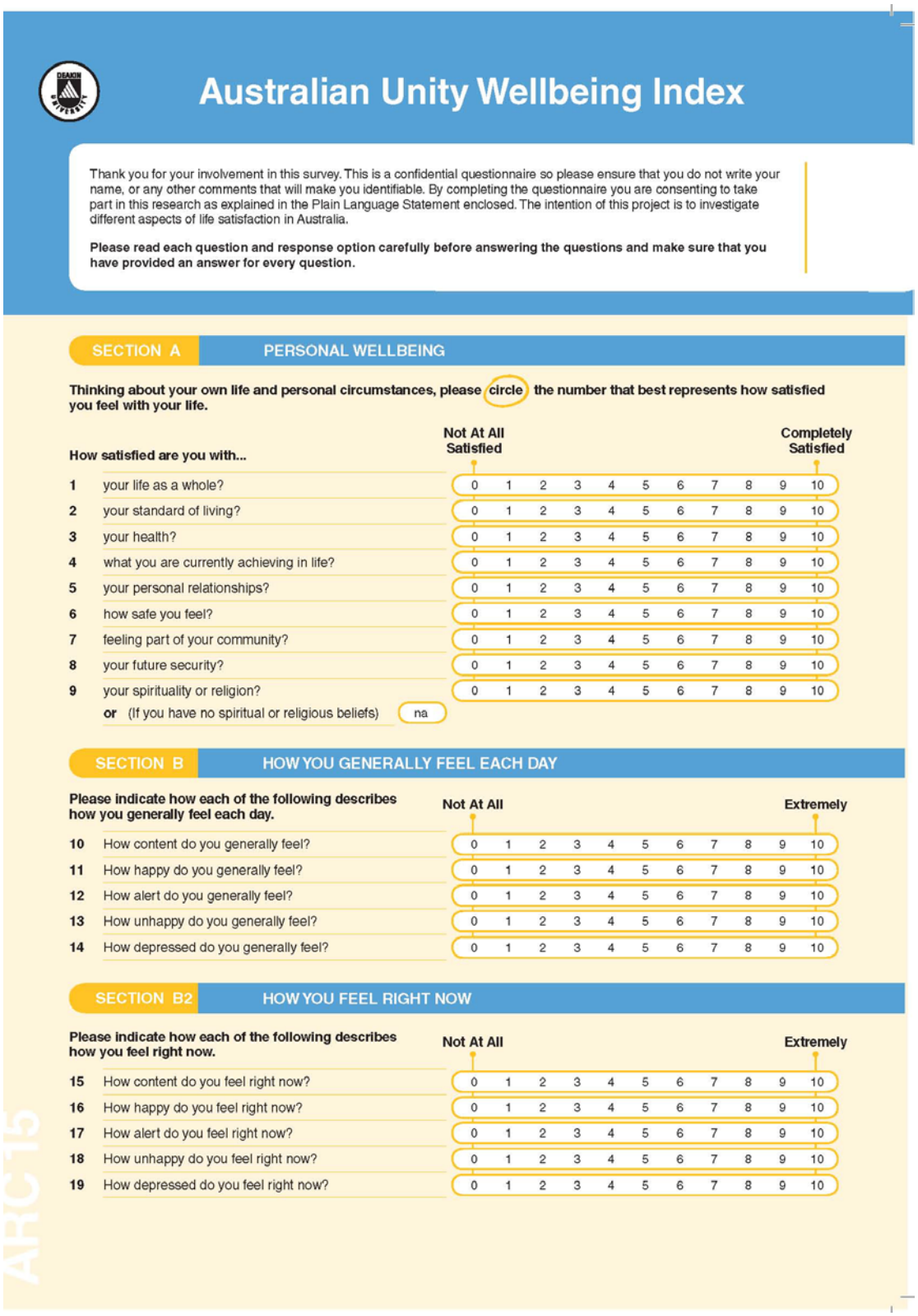
In addition you will be required to report on the progress of your project at least once every year and at the conclusion of the project. Failure to report as required will result in suspension of your approval to proceed with the project.

DU-HREC may need to audit this project as part of the requirements for monitoring set out in the *National Statement on Ethical Conduct in Human Research (2007)*.

V. Bates

Vicky Bates
On behalf of DU-HREC
(03) 9251 7052

Appendix 3: Copy of L15 Survey



Australian Unity Wellbeing Index

Thank you for your involvement in this survey. This is a confidential questionnaire so please ensure that you do not write your name, or any other comments that will make you identifiable. By completing the questionnaire you are consenting to take part in this research as explained in the Plain Language Statement enclosed. The intention of this project is to investigate different aspects of life satisfaction in Australia.

Please read each question and response option carefully before answering the questions and make sure that you have provided an answer for every question.

SECTION A — PERSONAL WELLBEING

Thinking about your own life and personal circumstances, please circle the number that best represents how satisfied you feel with your life.

How satisfied are you with...	Not At All Satisfied										Completely Satisfied
1 your life as a whole?	0	1	2	3	4	5	6	7	8	9	10
2 your standard of living?	0	1	2	3	4	5	6	7	8	9	10
3 your health?	0	1	2	3	4	5	6	7	8	9	10
4 what you are currently achieving in life?	0	1	2	3	4	5	6	7	8	9	10
5 your personal relationships?	0	1	2	3	4	5	6	7	8	9	10
6 how safe you feel?	0	1	2	3	4	5	6	7	8	9	10
7 feeling part of your community?	0	1	2	3	4	5	6	7	8	9	10
8 your future security?	0	1	2	3	4	5	6	7	8	9	10
9 your spirituality or religion?	0	1	2	3	4	5	6	7	8	9	10
or (If you have no spiritual or religious beliefs)	na										

SECTION B — HOW YOU GENERALLY FEEL EACH DAY

Please indicate how each of the following describes how you generally feel each day.

	Not At All										Extremely
10 How content do you generally feel?	0	1	2	3	4	5	6	7	8	9	10
11 How happy do you generally feel?	0	1	2	3	4	5	6	7	8	9	10
12 How alert do you generally feel?	0	1	2	3	4	5	6	7	8	9	10
13 How unhappy do you generally feel?	0	1	2	3	4	5	6	7	8	9	10
14 How depressed do you generally feel?	0	1	2	3	4	5	6	7	8	9	10

SECTION B2 — HOW YOU FEEL RIGHT NOW

Please indicate how each of the following describes how you feel right now.

	Not At All										Extremely
15 How content do you feel right now?	0	1	2	3	4	5	6	7	8	9	10
16 How happy do you feel right now?	0	1	2	3	4	5	6	7	8	9	10
17 How alert do you feel right now?	0	1	2	3	4	5	6	7	8	9	10
18 How unhappy do you feel right now?	0	1	2	3	4	5	6	7	8	9	10
19 How depressed do you feel right now?	0	1	2	3	4	5	6	7	8	9	10

ARC15

SECTION C OVER THE PAST WEEK

How much did these statements apply to you over the past week?

		Did Not Apply To Me At All	Applied To Me to Some Degree or, Some of the Time	Applied To Me to a Considerable Degree or, a Good Part of the Time	Applied To Me Very Much or, Most of the Time
20	I found it hard to wind down.	0	1	2	3
21	I couldn't seem to experience any positive feeling at all.	0	1	2	3
22	I found it difficult to work up the initiative to do things.	0	1	2	3
23	I tended to over-react to situations.	0	1	2	3
24	I felt that I was using a lot of nervous energy.	0	1	2	3
25	I felt that I had nothing to look forward to.	0	1	2	3
26	I found myself getting agitated.	0	1	2	3
27	I found it difficult to relax.	0	1	2	3
28	I felt down-hearted and blue.	0	1	2	3
29	I was intolerant of anything that kept me from getting on with what I was doing.	0	1	2	3
30	I was unable to become enthusiastic about anything	0	1	2	3
31	I felt I wasn't worth much as a person	0	1	2	3
32	I felt that I was rather touchy	0	1	2	3
33	I felt that life was meaningless	0	1	2	3

SECTION D EVENTS IN YOUR LIFE

34 Thinking back on your life, what is the highest level of happiness you have ever experienced?

No Happiness 0 1 2 3 4 5 6 7 8 9 10 Complete Happiness

35 Has anything happened to you recently causing you to feel happier or sadder than normal? Please tick as appropriate

○ Yes, happier ○ Yes, sadder ○ No—Please skip to Life in Australia (Item 38)

(If Yes) On a scale from 0 to 10, how strong would you rate this influence?

Very Weak 0 1 2 3 4 5 6 7 8 9 10 Very Strong

36 Which areas of your life have been strongly influenced by this event? Please tick **all areas** that have been affected.

○ Standard of living ○ Relationships ○ Achieving in life ○ Spiritual/Religious
○ Health ○ Personal safety ○ Connection to your community ○ Future security

37 Now please tick the **one single** life area that has been **most strongly affected.**

○ Standard of living ○ Relationships ○ Achieving in life ○ Spiritual/Religious
○ Health ○ Personal safety ○ Connection to your community ○ Future security

SECTION E LIFE IN AUSTRALIA

How satisfied are you with...

		Completely Dissatisfied				Neutral					Completely Satisfied
38	life in Australia?	0	1	2	3	4	5	6	7	8	9	10
39	the economic situation in Australia?	0	1	2	3	4	5	6	7	8	9	10
40	the state of the natural environment in Australia?	0	1	2	3	4	5	6	7	8	9	10
41	the social conditions in Australia?	0	1	2	3	4	5	6	7	8	9	10
42	government in Australia?	0	1	2	3	4	5	6	7	8	9	10
43	business in Australia?	0	1	2	3	4	5	6	7	8	9	10
44	national security in Australia?	0	1	2	3	4	5	6	7	8	9	10

SECTION F — OVER THE PAST TWO WEEKS

This section consists of 21 groups of statements. Please read each group of statements carefully, and then pick out the one statement in each group that best describes the way you have been feeling during the past two weeks, including today

Circle the number beside the statement you have picked. If several statements in the group seem to apply equally well, circle the highest number. Be sure that you do not choose more than one statement for any group, including Item 16 (Changes in Sleeping Patterns) or Item 18 (Changes in Appetite).

45 Sadness
- 0 I do not feel sad
- 1 I feel sad much of the time
- 2 I am sad all the time
- 3 I am so sad or unhappy that I can't stand it

46 Pessimism
- 0 I am not discouraged about my future
- 1 I feel more discouraged about my future than I used to be
- 2 I do not expect things to work out for me
- 3 I feel my future is hopeless and will only get worse

47 Past Failure
- 0 I do not feel like a failure
- 1 I have failed more than I should have
- 2 As I look back, I see a lot of failure
- 3 I feel I am a total failure as a person

48 Loss of Pleasure
- 0 I get as much pleasure as I ever did from the things I enjoy
- 1 I don't enjoy things as much as I used to
- 2 I get very little pleasure from the things I used to enjoy
- 3 I can't get any pleasure from the things I used to enjoy

49 Guilty Feelings
- 0 I don't feel particularly guilty
- 1 I feel guilty over many things I have done or should have done
- 2 I feel quite guilty most of the time
- 3 I feel guilty all of the time

50 Punishment Feelings
- 0 I don't feel I am being punished
- 1 I feel I may be punished
- 2 I expect to be punished
- 3 I feel I am being punished

51 Self-Dislike
- 0 I feel the same about myself as ever
- 1 I have lost confidence in myself
- 2 I am disappointed in myself
- 3 I dislike myself

52 Self-Criticalness
- 0 I don't criticise or blame myself more than usual
- 1 I am more critical of myself than I used to be
- 2 I criticise myself for all of my faults
- 3 I blame myself for everything bad that happens

53 Suicidal Thoughts or Wishes
- 0 I don't have any thoughts of killing myself
- 1 I have thoughts of killing myself, but I would not carry them out
- 2 I would like to kill myself
- 3 I would kill myself if I had the chance

54 Crying
- 0 I don't cry anymore than I used to
- 1 I cry more than I used to
- 2 I cry over every little thing
- 3 I feel like crying, but I can't

55 Agitation
- 0 I am no more restless or wound up than usual
- 1 I feel more wound up or restless than usual
- 2 I am so restless or wound up that it is hard to stay still
- 3 I am so restless or agitated that I have to keep moving or doing something

56 Loss of Interest
- 0 I have not lost interest in other people or activities
- 1 I am less interested in other people or things than before
- 2 I have lost most of my interest in other people or things
- 3 It's hard to get interested in anything

57 Indecisiveness
- 0 I make decisions about as well as ever
- 1 I find it more difficult to make decisions than usual
- 2 I have much greater difficulty in making decisions than I used to
- 3 I have trouble making any decisions

58 Worthlessness
- 0 I do not feel I am worthless
- 1 I don't consider myself as worthwhile and useful as I used to
- 2 I feel more worthless as compared to other people
- 3 I feel utterly worthless

59 Loss of Energy
- 0 I have as much energy as ever
- 1 I have less energy than I used to have
- 2 I don't have enough energy to do very much
- 3 I don't have enough energy to do anything

60 Changes in Sleeping Pattern
- 0 I have not experienced any change in my sleeping pattern
- 1a I sleep somewhat more than usual
- 1b I sleep somewhat less than usual
- 2a I sleep a lot more than usual
- 2b I sleep a lot less than usual
- 3a I sleep most of the day
- 3b I wake up 1-2 hours early and can't get back to sleep

61 Irritability
- 0 I am no more irritable than usual
- 1 I am more irritable than usual
- 2 I am much more irritable than usual
- 3 I am irritable all the time

62 Changes in Appetite
- 0 I have not experienced any change in my appetite
- 1a My appetite is somewhat less than usual
- 1b My appetite is somewhat greater than usual
- 2a My appetite is much less than usual
- 2b My appetite is much greater than usual
- 3a I have no appetite at all
- 3b I crave food all the time

63 Concentration Difficulty
- 0 I can concentrate as well as ever
- 1 I can't concentrate as well as usual
- 2 It's hard to keep my mind on anything for very long
- 3 I find I can't concentrate on anything

64 Tiredness or Fatigue
- 0 I am no more tired or fatigued than usual
- 1 I feel more tired or fatigue more easily than usual
- 2 I am too tired or fatigued to do a lot of the things I used to do
- 3 I am too tired or fatigued to do most of the things I used to do

65 Loss of Interest in Sex
- 0 I have not noticed any recent change in my interest in sex
- 1 I am less interested in sex than I used to be
- 2 I am much less interested in sex now
- 3 I have lost interest in sex completely

SECTION G — OVER THE PAST WEEK

Please respond to each of the following statements by ticking the appropriate box that indicates how you have been feeling over the past week.

66	I enjoy the things I used to enjoy	○ Definitely as much	○ Not quite so much	○ Only a little	○ Hardly at all
67	I can laugh and see the funny side of things	○ As much as I always could	○ Definitely not so much now	○ Not quite so much now	○ Not at all
68	I feel cheerful	○ Not at all	○ Not often	○ Sometimes	○ Most of the time
69	I feel as if I am slowed down	○ Nearly all the time	○ Very often	○ Sometimes	○ Not at all
70	I have lost interest in my appearance	○ Definitely	○ I may not take quite as much care	○ I don't take so much care as I should	○ I take just as much care as ever
71	I look forward with enjoyment to things	○ As much as I ever did	○ Rather less than I used to	○ Definitely less than I used to	○ Hardly at all
72	I can enjoy a good book or radio or TV programme	○ Often	○ Sometimes	○ Not often	○ Very seldom

SECTION H — THE KIND OF PERSON YOU ARE

How much do you agree with the following statements?

		Strongly Disagree					Neutral					Strongly Agree
73	I see myself as extraverted and enthusiastic.	0	1	2	3	4	5	6	7	8	9	10
74	I see myself as anxious and easily upset.	0	1	2	3	4	5	6	7	8	9	10
75	I see myself as reserved and quiet.	0	1	2	3	4	5	6	7	8	9	10
76	I see myself as calm and emotionally stable.	0	1	2	3	4	5	6	7	8	9	10

SECTION I — COPING WITH LIFE

How much do you agree that when something bad happens...

		Strongly Disagree					Neutral					Strongly Agree
77	I ask others for help or advice.	0	1	2	3	4	5	6	7	8	9	10
78	I look for different ways to improve the situation.	0	1	2	3	4	5	6	7	8	9	10
79	I use my skills to overcome the problem.	0	1	2	3	4	5	6	7	8	9	10
80	I remind myself that something good may come of it.	0	1	2	3	4	5	6	7	8	9	10
81	I remind myself that I am better off than some others.	0	1	2	3	4	5	6	7	8	9	10
82	I remember that the situation will improve if I am patient.	0	1	2	3	4	5	6	7	8	9	10
83	I don't do anything, as nothing can help.	0	1	2	3	4	5	6	7	8	9	10
84	I spend time by myself.	0	1	2	3	4	5	6	7	8	9	10
85	I just let my feelings out so others know how I feel.	0	1	2	3	4	5	6	7	8	9	10

SECTION J — MORE ABOUT PERSONAL WELLBEING

Thinking about your own life and personal circumstances, please circle the number that best represents how satisfied you feel with your life.

How satisfied are you with...

		Completely Dissatisfied					Neutral					Completely Satisfied
86	your life as a whole?	0	1	2	3	4	5	6	7	8	9	10
87	your standard of living?	0	1	2	3	4	5	6	7	8	9	10
88	your health?	0	1	2	3	4	5	6	7	8	9	10
89	what you are currently achieving in life?	0	1	2	3	4	5	6	7	8	9	10
90	your personal relationships?	0	1	2	3	4	5	6	7	8	9	10
91	how safe you feel?	0	1	2	3	4	5	6	7	8	9	10
92	feeling part of your community?	0	1	2	3	4	5	6	7	8	9	10
93	your future security?	0	1	2	3	4	5	6	7	8	9	10
94	your spirituality or religion?	0	1	2	3	4	5	6	7	8	9	10
	or (If you have no spiritual or religious beliefs)	na										

Thank you for your time and participation in this survey

Appendix 4: Plain Language Statement for New Longitudinal Participants

DEAKIN UNIVERSITY
PLAIN LANGUAGE STATEMENT

Dear
Last year you participated in the Australian Unity Well-being project that is conducted in conjunction with Deakin University. At that time, you indicated in a telephone survey that you would be willing to be involved in future surveys of this kind. We are writing to you now because we are conducting a longitudinal study of the well-being of Australians, to test a model of the factors that maintain well-being. We would like to invite you to be part of this study.

The research team involved is Associate Professor David Mellor, Professor Bob Cummins, Associate Professor Mark Stokes and Ms Renee Bear form Deakin University. Australian Unity is a partner in the project. Renee Bear will use part of this project for the purposes of her thesis in the Doctor of Psychology (Clinical) course at Deakin University.

Your participation in the project is voluntary. If you choose not to participate there are no adverse consequences. If you do agree to be involved, we will send you a questionnaire package once a year for the next three years. The questions will ask you to provide some basic demographic details, and then to answer some questions about yourself such as:

- How satisfied are you with your health?
- How satisfied are you with the common goals and values of people in your neighbourhood?

Other questions will ask you to indicate your level of agreement with various statements, such as:

- I found it difficult to work up the initiative to do things
- I often feel inferior to others.
- I really enjoy talking to people.
- When something gets in the way of a goal, I believe I can overcome it.
- I am able to do things as well as most other people.
- I'm always optimistic about my future.

You will be consenting to participate in this study by completing the enclosed questionnaire and returning it in the reply paid envelope provided. You may withdraw from the project at any time simply by not returning the questionnaire, or advising us that you do not wish to continue in the study. Should you withdraw, we will cease sending you questionnaires. However, as we will not be able to identify your responses to date if you withdraw, any responses you have provided prior to withdrawal will be used in the overall analysis.

In total, the questionnaire should take you about 25 minutes to complete, once each year. Your questionnaire will be given a code and your answers will be entered into a database for collation. The research team will not be able to identify you or your responses. The database will be securely stored electronically at Deakin University for 6 years, then it will be destroyed.

If for any reason you feel distressed by anything asked in the survey, we suggest that you contact Lifeline on 13 1114.

For further details of the study, please contact Associate Professor David Mellor on 03 9244 3742, or Professor Bob Cummins on 03 92446845.

Should you have any concerns about the conduct of this research project, please contact the Secretary, Ethics Committee, Research Services, Deakin University, 221 Burwood Highway, BURWOOD VIC 3125. Tel (03) 9251 7123 (International +61 3 9251 7123).

Appendix 5: Plain Language Statement for Continuing Longitudinal Participants

DEAKIN UNIVERSITY
PLAIN LANGUAGE STATEMENT

Dear Friend of the Australian Centre on Quality of Life,
Some time ago you participated in the Australian Unity Wellbeing project that is conducted in conjunction with Deakin University and investigating the well-being of Australians. At that time, you indicated that you would be willing to be involved in future surveys of this kind. We are writing to you now in the hope that you will wish to continue your active involvement with this project and thereby assist us in monitoring the wellbeing of Australians.

The research team involved is Associate Professor David Mellor, Professor Bob Cummins, Associate Professor Mark Stokes and Ms Renee Bear form Deakin University. Australian Unity is a partner in the project. Renee Bear will use part of this project for the purposes of her thesis in the Doctor of Psychology (Clinical) course at Deakin University.

Your participation in the project is voluntary. If you choose not to participate there are no adverse consequences. As for last year, participation in this longitudinal study will involve completing the enclosed questionnaire. The questionnaire asks you to provide some basic demographic details, and to answer some questions about how you feel such as:

- How satisfied are you with your health?
- How satisfied are you with the common goals and values of people in your neighbourhood?

Other questions will ask you to indicate your level of agreement with various statements, such as:

- I found it difficult to work up the initiative to do things
- I really enjoy talking to people.
- When something gets in the way of a goal, I believe I can overcome it.

We remind you that your participation in the project is voluntary. You will be consenting to participate in this study by completing the questionnaire and returning it in the reply paid envelope provided. You may withdraw from the project at any time simply by not returning the questionnaire, or by advising us that you do not wish to continue in the study. Should you withdraw, we will cease sending you questionnaires. However, as we will not be able to identify your responses to date if you withdraw, any responses you have provided prior to withdrawal will be used in the overall analysis.

The questionnaire should take about 25 minutes to complete. We have included a reply paid envelope to return your completed questionnaire to Deakin University, and when you do this, we will assume you are doing so willingly. Your questionnaire is coded by number and the research team will not be able to identify you or your personal responses. However, we will enter your answers into a database which will allow us to collate your responses with your previous survey data. Because the questionnaire is coded and you do not record your name on it, the research team members are not able to identify the responses from any individual. In this way your confidentiality is maintained. The original database will be securely stored electronically at Deakin University for 6 years and then destroyed. However, the extracted and coded data will be retained for research purposes.

If for any reason you feel distressed by anything asked in the survey, we suggest that you contact Lifeline on 13 1114.

For further details of the study, please contact Professor Bob Cummins on 03 92446845 or Associate Professor David Mellor on 03 9244 3742.

Should you have any concerns about the conduct of this research project, please contact the Secretary, Ethics Committee, Research Services, Deakin University, 221 Burwood Highway, BURWOOD VIC 3125. Tel (03) 9251 7123 (International +61 3 9251 7123).