VALIDATION OF THE BDI-II IN SOUTH AFRICA

by

MAKHUBELA MALOSE SILAS

THESIS

Submitted in fulfilment of the requirement for the degree of

DOCTOR OF PHILOSOPHY

In

PSYCHOLOGY

In the

FACULTY OF HUMANITIES

(School of Social Sciences)

at the

UNIVERSITY OF LIMPOPO

SUPERVISOR: PROF. S MASHEGOANE

2015

DECLARATION

I declare that the thesis hereby submitted to the University of Limpopo, for the degree of Doctor of Philosophy in Psychology has not previously been submitted by me for a degree at this or any other university; that it is my work in design and in execution, and that all material contained herein has been duly acknowledged.

____________________________ ____________________

Initials & Surname **Date**

ABSTRACT

The present study investigated whether the Beck Depression Inventory-II (BDI-II) measures the same construct in exactly the same way across the groups of interest and time in South Africa. The degree to which items or subtests of the BDI-II have equal meaning across qualitatively distinct groups of examinees (e.g., culture and gender) was explored. Measurement Invariance (MI) of the BDI-II across race (blacks and whites), gender and time (two weeks lag) was examined in a sample of university students, from two universities located in diverse geographical regions of South Africa (N = 919). Confirmatory factor analysis (CFA) was used to test the fit of the hypothesized three-factor model established through exploratory factor analysis (EFA), and the results from these analyses indicated that the BDI-II was most adequately represented by a three lower-order factor structure (appropriately named Negative attitude, Performance difficulty and Somatic complaints). Results based on multigroup confirmatory factor analysis (MCFA) (i.e., means and covariance structures [MACS]) indicated that there was factorial invariance for this three lower-order factor structure across groups and time, suggesting that the BDI-II provides an assessment of severity of depressive symptoms that is equivalent across race, gender and time in university students. Results indicated that MI was established at the level of configural, metric and scalar invariance for race, gender and across time. However, there was some evidence of differential item functioning (DIF) and differential additive response style (ARS) across race, with two noninvariant intercepts (items 5 and 14) and three item intercepts (items 11, 14 and 18) across gender being identified. Additionally, results of latent mean differences were presented to explain group differences. The study concluded with recommendations for future studies.

DEDICATION

This thesis is dedicated to my family for their support and encouragement.

ACKNOWLEDGEMENTS

I would like to thank the following people:

Professor S. Mashegoane, my supervisor, who mentored me throughout this testing process. As Isaac Newton once intimated: ***"if I've seen further, it's because I'm standing on the shoulders of giants"***.

Students from both the University of Limpopo and the University of Pretoria, who participated in the study; and

Lecturers from both institutions who gave access to their students to participate in the study.

The research assistants who helped with data collection.

Dr. L.K Debusho (UP), who provided counsel on certain aspects of statistical analysis.

The National Research Foundation (NRF) for their financial support that enabled the completion of this study.

TABLE OF CONTENTS

Chapter 5: Discussion, Recommendations and Limitations of the study

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	Affective factor
ARS	Additive response style
B-B NNFI	Bentler-Bonett Non-Normed Fit Index
BDI-II	Beck Depression Inventory II
BHS	Beck Hopelessness Scale
C	Cognitive factor
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
COVS	Covariance structures
df	Degrees of freedom
DALY	Disability adjusted life years
DIF	Differential Item Functioning
DSM IV	Fourth edition of the Diagnostic and Statistical Manual of Mental disorders
DSM V	Fifth edition of the Diagnostic and Statistical Manual of Mental disorders
EFA	Exploratory Factor Analysis
”g”	General factor
HSCL-25	Hopkins Symptom Checklist-25
IRT	Item Response Theory
KMO	Kaiser-Meyer-Olkin measure of sample adequacy
LM	Lagrange Multiplier
MACS	Mean and covariance structures
MCFA	Multigroup confirmatory factor analysis
MI	Measurement Invariance
ML	Maximum Likelihood
MTMM	Multitrait-Multimethod Matrix
NA	Negative attitude
PCA	Principal Component Analysis
PD	Performance difficulty
PSS-10	Perceived Stress Scale-Ten Item Version
RMSEA	Root mean square error of approximation

RSES	Rosenberg Self-Esteem Scale
S	Somatic factor
S-B χ^2	Satorra-Bentler scaled Chi-square test
SCID-I/NP	Structured Clinical Interview-I/Non-patient version
SEM	Structural Equation Modeling
SGSES	Sherer General Self-Efficacy Scale
SRMR	Standard root mean squared residual
Y-B χ^2	Yuan-Bentler scaled Chi-square test
z	Wald significance test

CHAPTER 1
OVERVIEW OF THE STUDY

1. General introduction

1.1 Introduction

According to the World Health Organisation (2001), unipolar depressive disorders are the foremost cause of disability globally and rank fourth in the ten prominent causes of the global burden of disease (Murry & Lopez, 1996). In the 2001 census, Statistics South Africa reported that 16% of the South African population (with a population of approximately 44 500 000 people, at the time) suffered from mood disorders (Statistics South Africa, 2005). Tomlinson, Grimsrud, Stein, Williams and Myer (2009) found in their epidemiological study the incidence of major depression in South Africa to be 9.8% for lifetime and 4.9% for the past 12 months (i.e., relatively higher than Nigeria and China). They further established that over 90% of all respondents with depression in their study reported global role impairment.

In his inaugural lecture as professor in the Department of Psychology at the then University of the North, Peltzer indicated that:

> "In the year 2020 it is calculated that the single highest cause of Disability Adjusted Life Years (DALY) will be depression (5.6%). In addition, because suicide is a far greater risk in depression, depression has a greater impact on premature mortality" (Peltzer & Habil, 1998, p. 4).

With a high reported prevalence rate (e.g., 9.8% lifetime depression) of depression and its ensuing consequences to the society (i.e., disability and premature mortality), it is imperative that significant levels of depressive symptoms be recognized, assessed, and treated. The utility and psychometric properties of the Beck Depression Inventory–II (BDI–II; Beck, Steer, & Brown, 1996), considered to be the gold standard for identifying depression in adolescents and adults (Camara, Nathan, & Puente, 2000), are well established in Western countries and remain to be determined in South Africa.

The BDI–II is one of the most frequently used measures of the severity of depression in adolescents and adults by both researchers and clinicians (Brantley, Dutton, & Wood, 2004). Its items encompass the diagnostic criteria for major depression of the fourth edition of the Diagnostic and Statistical Manual of Mental disorders (DSM-IV, American Psychiatric Association, 1994). Studies have reported the factor structure of the BDI–II in different populations with varying consistency (Arnau, Meagher, Norris, & Bramson, 2001; Beck et al., 1996; Steer et al., 1999). Although these studies indicate that the BDI–II has moderately stable psychometric properties across groups, the majority of subjects were Western and middle-class, with no study including more than 15% of their sample as African American or non-Western subjects. As a result, little is known about its psychometric properties in non-Western samples.

1.2 Statement of the problem

The validity of clinical psychological measures remains indeterminate; particularly when tests are applied in cultural groups other than those for which they were originally developed (Swanepoel & Kruger, 2011). Implicit in this is the underlying assumption that most assessment measures do not measure the same constructs in exactly the same way across the groups of interest; that is, the instruments are not measurement equivalent. Research evidence confirms that cultural differences affect responses to certain psychological measures and reduce their validity for specific cultural groups (de Klerk, Boshoff & van Wyk, 2009; Meiring, Van de Vijver, & Rothmann, 2006; Van Eeden & Mantsha, 2007). The mainstream of standardized tests used in Africa have been imported from Western countries because of the inadequacy of psychological research and development of culturally appropriate, reliable and valid psychological measures. Several of these tests have simply been transferred and applied without considering the possibility of test bias, and no attempt was made to adapt and/or create norms for them.

Although some studies find support for the validity of psychometric instruments developed elsewhere when applied to South African samples (e.g., De Bruin, Swartz, Tomlinson, Cooper, & Molteno, 2004; Storm & Rothmann, 2003), most studies indicate validity, reliability and structural problems when such instruments are applied within the South African context without adaptations (e.g., Abrahams & Mauer, 1999; Gray & Durrheim, 2006; Meiring et al., 2006; Van Wyk, Boshoff & Owen, 1999). The diverse population of South Africa confounds this problem even more. Research has found that some scales demonstrate adequate validity when applied to native English and/or Afrikaans speaking South Africans compared to other cultural/ethnic groups (e.g., Abrahams, 2002; Abrahams & Mauer, 1999; Claassen, 1997; Owen, 1992; Rushton, Skuy, & Fridjhon, 2003; Tyler, 2000). Results from these studies confirm that it is risky to apply an instrument developed in the USA or elsewhere to South Africa without revalidating the instrument. Psychological constructs can be perceived distinctively across cultures.

The BDI–II is a commonly used measure of depression by clinicians in South Africa, aiding with the diagnosis, case formulation and the monitoring of response to treatment for depression (Steele & Edwards, 2008). Yet, the Human Sciences Research Council (2007) in South Africa reported a lack of African norms, psychometric properties and published or unpublished validation studies. There is however two published and three unpublished African research studies using the BDI-II (Drennan, 1988; Lester & Akande, 1995; Pillay & Sargent, 1999; Steele, 1996; Westaway & Wolmarans, 1992). Steele and Edwards (2008) developed and validated Xhosa translations of the Beck scales, but the process of translation failed to yield trustworthy results (i.e., translation equivalence). However, translated instruments, regardless of their reconstructive precision, address only the linguistic equivalence of an assessment measure which is inadequate (see, e.g., Byrne, Stewart, & Lee, 2004; Cheung, Leon, & Ben-Porath, 2003; Leung & Wong, 2003; van de Vijver & Hambleton, 1996). Beyond this rudimentary requirement lies the need to establish their construct validity and psychometric data before claims of complete adaptation for use with a particular population (e.g., South African) can be made.

It would appear that only limited validation studies have been conducted on the BDI–II in Africa. Accordingly, further studies are necessary to determine its usefulness in the African context.

1.3 Background of the study

The assessment of latent constructs, such as psychological states (e.g., depression) serves an indispensable role in society, particularly when test scores purporting to measure these constructs are used to inform decision making in a variety of settings (French & Finch, 2006). A significant upsurge in the use of test scores for such decisions has been observed across clinical, educational, forensic and occupational settings (Brennan, 2004). Therefore, the statistical properties of tests ought to meet validity standards to overcome both legal and technical challenges from the clinical community and the general public.

The evaluation of measurement invariance (MI; i.e., the degree to which items or subtests have equal meaning across groups of examinees) is one method used to gather score validity evidence and to appraise construct-irrelevant variance (e.g., group affiliation). The examinee’s score should not depend on construct-irrelevant variance. The possibility of committing serious errors is high, if decisions are made for individuals in the absence of MI (Bollen, 1989). Observed score differences can either reflect true group mean differences or differences in the relation between the construct and the observed score that is not equivalent across groups (Raju, Laffitte, & Byrne, 2002). Thus, to circumvent undesirable social consequences, the measurement process must keep irrelevant variables from influencing scores and employ methods to determine the extent to which scores are influenced by such variables (Messick, 1989).

Various degrees of MI have been defined (e.g., Little, 1997; Meredith, 1993; Millsap, 2005). For example, Little (1997) proposed two hierarchical levels of MI. The first level requires that the psychometric properties of an instrument be equivalent (i.e., configural, metric, measurement error, and scalar invariance) and the second level comprises of group differences in latent means and covariances (French & Finch, 2006). These levels of MI are examined with a MCFA, which allows for testing a priori theory of the test structure across groups or across time (i.e., developmentally related questions) (Mantzicopoulos, French, & Maller, 2004). This approach allows for the comparison of specific features of the factor model from one group to another. When these features are found to be equivalent across groups, MI (specifically, factorial) can be inferred.

1.4 Aim of the study

The study investigates the validity of the BDI-II among South African university students.

1.5 Objectives of the study

To answer the research questions, the following objectives were developed:

1.5.1 To evaluate if there will be a conceptually meaningful factor structure of the BDI-II, using both EFA and confirmatory factor analysis CFA followed by cross-validation of the determined factor structure using the CFA;

1.5.2 To assess the factorial invariance on the BDI-II across gender and race, determining the generalizability of psychometric properties across male and female, and white and black South African samples;

1.5.3 To assess the MI of the resulting model longitudinally across two points; and

1.5.4 To examine if there will be empirically and theoretically justifiable correlations between depression and related external criteria.

1.6 Research questions:

This study is designed to answer the following research questions:

1.6.1 Will the BDI-II yield a conceptually meaningful and parsimonious factor structure with both EFA and CFA in South Africa?

1.6.2 Will there be any factorial invariance of the resulting model between identified demographic subgroups (i.e., gender and race)?

1.6.3 Will there be longitudinal MI of the BDI-II?

1.6.4 Will depression measured with the BDI-II have empirically and theoretically justifiable correlations with external criteria (i.e., hopelessness, perceived stress, self-efficacy and self-esteem)?

1.7 Contribution of the study to theory and practice

The measurement of psychopathology in cross-cultural milieus has often had serious limitations (Wang, Andrade, & Gorenstein, 2005). This may be particularly problematic regarding self-reporting assessment measures for mood disorders, where the understanding of the meaning of affect-loaded items relies on the respondent's interpretation. Reporting depressive symptoms may also be influenced by serious cultural biases in non-Western populations (e.g., language and social desirability of some behaviours), thereby resulting in poor validity (Wang et al., 2005). Eurocentric psychological knowledge, advances and experiences are often adopted without any validation by researchers in different countries resulting in unforeseen perils, and for this reason, validating psychological assessment measures with non-Western populations is an imperative endeavour. The investigation will contribute to the utility and application of the BDI-II in non-western countries, specifically in South Africa.

1.8 Operational definition of terms

1.8.1 Depression:

Depression in the context of this study will refer to a clinical syndrome, or cluster of symptoms, covering changes in affect, cognition and behaviour, and which meet the diagnostic criteria for a Major Depressive Disorder according to the DSM 5 (American Psychiatric Association, 2013; Beck, 1967). Clinical depression is characterised by a pervasive sad mood, anxiousness, hopeless, rumination, anhedonia, suicidal ideation, tearfulness, and vegetative and somatic complains; whereas non-clinical depression/dysphoria is a normal reaction to certain life events or stressful events and not always a psychiatric disorder.

1.8.2 Validation:

Validation generally refers to the evaluation of a psychological measure for its psychometric properties and suitability as a measuring instrument, before it is relied on for making decisions. Psychometric validation of a measure, in this instance, includes the evaluation of aspects like construct validity across populations and contexts, MI across different populations, reliability of the measure and stability over time, and the general responsiveness of the test using various statistical methods. This involves an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores (Messick, 1989). This process also includes consideration for test relevance and utility as well as more traditional validity measures relating to content, criteria, and construct models and it is used likewise in the current study.

1.8.3 MI:

MI, a method used to gather validity evidence, is the degree to which items or subtests have equal meaning across qualitatively distinct groups of examinees (e.g., age, gender and culture). It includes methods like configural, metric, measurement error, and scalar invariance. MI with respect to groups is an essential aspect of the fair use of scores of psychological measurements. It can tell us whether the same factor structure upholds across groups, and also whether the factor loadings and the means and error distribution of the groups are the same. By testing for MI, we test for equivalence of the structural equation model across groups. It is used similarly in the current study.

1.8.4 EFA:

EFA could be described as an orderly simplification of interrelated measures. It further examines and explores the interdependence among the observed variables in some set. It has traditionally been used to explore the possible underlying factor structure of a set of observed variables without imposing a preconceived structure on the outcome. By performing EFA, the number of constructs and the underlying factor structure are identified. EFA is used in the present study to help develop a structural theory: to determine the number of latent constructs/factors underlying a set of items (variables), select "best" measures of a construct and to define the content or meaning of factors (latent constructs).

1.8.5 CFA:

CFA is a statistical technique used to verify the factor structure of a set of observed variables. It allows the researcher to test the hypothesis that a relationship between observed variables and their underlying latent constructs exists. The researcher uses knowledge of the theory, empirical research, or both, postulates the relationship pattern a priori and then tests the hypothesis statistically. Likewise in the current study, CFA is used to test hypotheses corresponding to prior theoretical notions, which include the number and nature of factors, and more complex hypotheses, such as the equality of factor pattern matrices across groups.

1.9 Conclusion

This section outlined the research questions, aim and objectives of the study and further introduced concepts relevant to the study.

CHAPTER 2
THEORETICAL PERSPECTIVE AND LITERATURE REVIEW

2. Introduction

This chapter outlines the theoretical framework that frames all the expositions and discussions in the study. Finally, the literature on the BDI-II's factor structure across contexts, populations, culture and gender is reviewed.

2. 1 THEORETICAL PERSPECTIVE: MESSICK'S CONTEMPORARY THEORY OF UNIFIED VALIDITY

Validity theory has evolved significantly over the last century in response to the increased use of assessments across scientific, clinical, educational, occupational and legal settings (Anastasi, 1986; Kane, 2006; Messick, 1989). The primary trajectory of this evolution reveals a shift to a notion of validity reliant on the interpretation of multiple evidence sources and the recognition that validity cannot be captured by one single score (DeLuca, Klinger, Searle, & Shulha, 2010; Kane, 2001; Messick, 1989; Moss, 1998). Contemporary validity theory and practice recognizes that "since predictive, concurrent, and content validities were all essentially ad hoc, construct validity was the whole of validity from a scientific point of view" (Loevigner, 1957, p. 636). Messick's (1989) unitary view of validity that involves the integration of both evidential and consequential sources of evidence (i.e., *Facets of Validity Matrix*) is such a theory, and it is used in the present study as the theoretical foundation for the examination of measurement validity of the BDI-II.

2.1.1 Messick's contemporary theory

Drawing on the work of Cronbach (1984; 1988), Messick describes validity as "an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (Messick, 1989, p. 13). Messick's (1989) theory of unified validity is a synthesis of competing validity theories and corresponding validation techniques (Crocker & Algina, 1986). This unitary framework integrates both evidential (i.e., factual) and consequential sources of evidence of validity. The practice of validation aims to ascertain the extent to which an interpretation of a test is conceptually and empirically warranted and should be aimed at making explicit any ethical and social values that overtly or inadvertently influences that process (Anastasia, 1986; Cronbach, 1984; DeLuca et al., 2010; Guion, 1977; Messick, 1989; Zumbo, 2009). Evidential validity includes consideration for test relevance and utility as well as more traditional validity measures relating to content, criteria, and construct models. Consequential validity considers the ramifications of test interpretations and use on various stakeholders and includes value implications and social consequences (DeLuca et al., 2010).

A critical distinction between the obsolete ideas of “type of validity” (face, construct, criterion, etc.) and the unitary concept of validity is the focus on seeking “sources of validity evidence” as opposed to identifying types of validity. This notion applies aptly to both paper-and-pencil based assessments (e.g., BDI-II) and performance based assessments. Messick’s (1989) contemporary theory of validity is the essence of the approach for examining score validity in common use currently. The model is grounded in the seeking of evidence of score interpretation validity from six possible sources. These sources are content, substantive, structural (internal), external, generalizability and consequential (Crocker, 1997; Messick, 1992). The present study explores evidence of the BDI-II score validity within unitary validity theory by construct validity measurements including: (a) correlations between a measure of the construct and the designated construct theory, (b) mean score differentiation between groups, (c) factor analysis, and (d) multitrait-multimethod matrix (MTMM).

Content as a source of construct validity: Content validity evidence resides in the relation between the test and the domain of application. It considers how well the respondent’s responses to an assessment reflect their knowledge of the content area (DeLuca et al., 2010). The content aspect of validity comprises identifying the boundaries of the domain to be assessed, then assuring that the tasks and questions chosen for the assessment reflect those boundaries and are relevant to and representative of the domain (Kane, 2006). If an area of the domain is not represented, there is possible risk of construct under-representation; if a quality is being measured other than what is anticipated, there is a threat of construct-irrelevant variance. In either instance the score interpretation will not be considered valid for its intended use. The present study considers content as a source of construct validity by exploring the possible domains being measured in the BDI-II.

Substantive sources of construct validity: With substantive sources of construct validity evidence, the content aspect is expanded upon to accommodate a requirement of empirical evidence (DeLuca et al., 2010; Kane, 2006). Items for the ultimate assessment instrument are chosen based on consistent empirical responses to the tasks. The evidence moves beyond expert judgement of content appropriateness by employing structural equation modeling (SEM) techniques, such as factor analysis. In the case of factor analysis, how the task loads on a factor provides support for the domain representation and validity of score interpretation. The present study examines substantive sources of construct validity through a CFA of the refined BDI-II subscales.

Structural (internal) sources of construct validity: It addresses the relations among responses to the tasks, items, or parts of the test. The internal structure of the test should be consistent with the internal structure of the domain (Cronbach & Meehl, 1955; Kane, 2006). An item analysis, for instance, may reveal that respondents with high overall scores all answered a particular item incorrectly. Similarly, respondents with low overall scores all answered that same item correctly. This should prompt an investigation of the item, as the relationship between the item response and the test responses as a whole is unexpected. The present study seeks evidence of the BDI-II score validity by exploring the internal structure of the BDI-II at the item level using SEM (i.e., EFA and CFA).

External sources of construct validity: It explores the relationships between the assessment scores of interest and other measures and related variables of either the same or different constructs (Cronbach & Meehl, 1955; Kane, 2006). Convergent evidence might be identified through high correlations between scores from tests of a similar construct, while discriminant evidence might be identified through low correlations with scores from tests of unrelated constructs. The present study seeks evidence of BDI-II score validity by exploring the relationships between the scores on the BDI-II and scores on depression-related measurement of constructs using MTMM approach (convergent and divergent validity).

Generalizability as a source of construct validity: Generalizability refers to the variances in the test processes and structures over time, across groups and settings, or in response to experimental interventions (BDI-II MI). A narrower set of tasks may provide better reliability, but at the expense of generalizability across populations or settings. Equally, a broader set of tasks may provide more generalizability, but at the expense of reliability. Generalizability applies to the transfer of tasks across the domain; that is, how does performance on one task predict performance on another. Configural, metric and scalar invariance of the BDI-II will be explored across gender, race and over time.

There is, however, no set formula prescribing how many sources or how evidence from each source is sufficient to substantiate a validity claim. Messick (1989) states that test score validity does not depend on any particular source of validity, nor does it require any one particular source. Rather, one looks for a compelling argument that uses convergent and discriminant evidence to support the score interpretation. A review of the literature on the BDI-II reveals a focus on the structural, substantive evidence and external aspects of construct validity (i.e., explored through its factor structure and correlation of BDI-II scores with other assessment scores).

2.2 LITERATURE REVIEW

2.2.1 Beck Depression Inventory (BDI)

The BDI is one of the most widely used depression self-rating scales. More than 2,000 empirical studies have employed it (Richter et al., 1998). Although originally developed as a measure for clinical populations, it is now also widely used as a measure of depression in non-clinical/subclinical samples. The BDI has subsequently been upgraded to the BDI II (Beck, Steer, & Brown, 1996) to make its symptom contents more comparable to the diagnostic criteria for major depressive episode in the DSM IV (American Psychological Association, 1994).

Beck et al. (1961) developed the BDI as a self-report measure to assess the behavioural manifestations of depression. It was empirically constructed based on a pool of pre-selected items from Beck's observations in psychotherapy with depressed patients (Steer et al., 1986). These symptoms (i.e., 21 items) include disturbed mood (sadness, loneliness, apathy) and a negative self-concept and self-punitive wishes. Further, depressive persons can suffer from somatic/vegetative symptoms such as anorexia, insomnia, and loss of libido. Depressed individuals can also experience changes in activity level: they show either retardation or agitation (Beck, 1967).

2.2.2 Factor structure of the BDI II across different contexts

While the BDI-II is often used with diverse populations (culturally, racially and geographically), limited empirical support for its factor solution exists outside of North America. Although aspects of the reliability and validity of the BDI-II have been supported by research, its factor structure has varied across studies (Arnau et al., 2001; Beck et al., 1996; Vanheule, Desmet, Groenvynck, Rosseel, & Fontaine, 2008). A number of studies have been interpreted to support two factors, which roughly reflect cognitive symptoms and somatic symptoms, whereas other research has supported a three-factor model (see Shafer, 2006). Items in the BDI-II reflect Cognitive (C), Affective (A), and Somatic (S) components of depression (Beck, 1996). In methodological studies, these three components are regularly represented in factor analytic findings.

2.2.2.1 **Two-factor structure**

The BDI-II was originally validated using an outpatient sample (N = 500) and an undergraduate sample (N = 120) (Beck et al., 1996). Both samples yielded two factors in EFA, using items that loaded ≥ 0.35 on the corresponding factors. The factors for the outpatient sample were labeled "Somatic-Affective" (SA) and "Cognitive" (C) (the so-called SA-C model). The factors for the undergraduate sample were labeled "Cognitive-Affective" (CA) and "somatic" (S) (i.e., CA-S model). In subsequent CFA studies using all the items of the BDI-II, these two-factor models were confirmed respectively for a clinically depressed outpatient group (Steer, Ball, & Ranieri, 1999) and for samples of undergraduate students (Storch, Roberti, & Roth, 2004; Whisman, Perez, & Ramel, 2000).

The first two-factor model reported by Beck et al. (1996) consisted of correlated C-SA factors consisting of 9 and 12 items, respectively. Revision of the 21-item inventory revealed 19 of the 21 items loaded on a two-factor solution (Beck et al., 1996) (SA items "pessimism" and "loss of interest in sex" had loadings less than 0.35). In contrast, Whisman et al. (2000) applied CFA to test the CA-S model derived from the Beck et al. (1996) study. They found that 20 of the items loaded (> 0.35) on either a CA or S factor solution in a sample of 576 college students. However, the "loss of interest in sex" item had a loading of 0.31 on S.

Using a sample of 414 North American university students, Storch et al. (2004) reported a two-factor solution (CA and S) that included all items, with items assessing "pessimism" and "loss of interest in sex" loading on the CA factor, derived from CFA. Furthermore, researchers have like Roth (2004) also reported a two-factor structure, with C and A symptoms loading onto one factor and S symptoms loading onto a second factor ("CA-S" structure: 16 items and 5 items on the first and second factor respectively) (Beck et al.,1996; Buckley, Parker, & Heggie, 2001) while others have reported S and A items loading onto one factor and C items loading onto a second factor

("SA-C" structure) (Arnau et al., 2001; Beck et al., 1996; Gary & Yarandi, 2004; Steer et al.,1999). The BDI–II's factor structure has been examined in different populations with this persisting inconsistency (Grothe et al., 2005).

A two-factor solution (factors: C & S) has been identified for psychiatric outpatients (Beck et al., 1996), primary care medical patients (Arnau et al., 2001), clinically depressed outpatients (Steer et al., 1999), depressed geriatric inpatients (Steer, Rissmiller, & Beck, 2000), college students (Beck et al., 1996; Dozois, Dobson, & Ahnberg, 1998; Whisman et al., 2000) and non-clinical high school adolescents (Osman, Barrios, Gutierrez, Williams, & Bailey, 2008). Steer and Clark (1997) applied EFA to data from 160 students and observed a factor structure that fit the CA–S pattern except that "agitation" loaded on the S factor and "irritability" had a substantial cross-loading on the S factor. Dozois et al. (1998) formulated an alternative two-factor structure in a student sample with a CA (10 items) factor and a Somatic-Vegetative (SV) (11 items) factor.

Arnau et al. (2001) formulated a two-factor model using data collected in a primary care medical setting and found a SA factor (12 items) and a C factor (8 items). This model contains one item (Item 8) with high loadings on both factors, which implies that the model is not unidimensional. Penley, Wiebe and Nwosu (2003) replicated this model (CFA) in a sample of hemodialysis patients (n = 122). Kojima et al. (2002) also confirmed the SA (12 items) and C (9 items) factors model by means of CFA in a general population (n = 353). Wiebe and Penley (2005) also observed good fit for the SA-C model with CFA, using student samples (respectively: n = 160, n = 895).

Various studies have conducted a CFA of the BDI in clinical populations (Dunkel, Froehlich, Antretter, & Haring, 2002; Johnson, DeLuca, & Natelson, 1996; Miles et al., 2001; Morley, Williams, & Black, 2002), and a general two-factor model also emerged from these studies (i.e., BDI-II items 1-14: CA symptoms and items 15-21: S or V symptoms). Other studies with cardiac patients have conducted a principal component analysis (PCA) of BDI-II data to derive two factors with cross loading items (SA and CA) and a third questionable two-item appetite factor (de Jonge et al., 2006; Linke et al., 2009; Martens et al., 2010). Viljoen, Grant, Griffiths and Woodward (2003) found the BDI to have a two-factor solution, with a sample of medical outpatients, which they referred to as SA (10 items) and C (9 items). In this model, two items (Item 21; Item 13) loaded on both factors, and one item (Item 18) does not have adequate loadings on any of the factors. Using a sample of psychiatric inpatients, Cole, Grossman, Prilliman and Hunsaker (2003) found two scales (C and S) that were similar to that reported for medical outpatients (Viljoen et al., 2003).

Osman, Kopper, Barrios, Gutierrez and Bagge (2004) conducted CFA of BDI-II data obtained from adolescent boys and girls residing in inpatient psychiatric units, testing two-factor and three-factor models and found that none of them fitted well with the data. However, a subsequent EFA revealed a two-factor solution (CA and SA) but one that differed considerably from the two-factor solutions previously reported. With a sample of clinically depressed outpatients, Steer et al. (1999) found a two-factor solution and one second-order depression factor. This model contains a C factor (8 items) and a factor that was this time labeled Noncognitive (13 items). Segal, Coolidge, Cahill and O'Riley (2008) found a two-factor structure of the BDI–II with a sample of community-dwelling older and younger adults. Kneipp, Kairalla, Stacciarini and Pereira (2009) also recently found a two-factor structure among low-income women with chronic health conditions, with the cognitive and affective domains represented in Factor 1, and somatic items comprising Factor 2. Vanheule et al.'s (2008) study further supported Beck's original model (SA and C), with a good fitting structure containing 15 and 16 items developed with an item-deletion algorithm. Wu (2010) also identified a two-factor solution (CA and S) in high school students in Taiwan.

2.2.2.2 **Three-factor structure**

Several authors identified a three-factor solution with adolescent psychiatric outpatients and college students (i.e., C, SA, and guilty feelings and punishment feelings) (Carmody, 2005; Osman et al., 1997; Steer et al., 1998;

Vanheule et al., 2008). This is consistent with the already identified hierarchical structure, consisting of three first-order factors and one second-order factor structure of the BDI-II with high school and college students (e.g., Byrne & Stewart, 2006; Byrne, Stewart, Kennard, & Lee, 2007; Johnson, Neal, Brems, & Fisher, 2006; Rowland, Lam, & Leahy, 2005; Whisman, Judd, Whiteford, & Gelhorn, 2012). Carmody's (2005) CFA study of the BDI-II provided comparative data with an ethnically diverse university sample (n = 502) demonstrating a robust fit for the three-factor model (Negative attitude: NA, Performance difficulty: PD, and Somatic symptoms: S).

Osman et al. (1997) found a three factor model based on a BDI factor model developed by Byrne and Baron (1993). Through CFA, the authors observed an adequate model fit with student data. The factors of this model are NA (10 items), PD (7 items), and S (5 items). The model contains three correlations between residual variances and one item (Item 10) that loads on two factors. Beck, Steer, Brown and van der Does (2002) using a clinical and nonclinical data also found a three-factor model by applying simultaneous component analysis with promax rotation. The authors obtained a model with C (7 items), S (9 items), and A (5 items) factors that was judged to be theoretically plausible. Buckley et al. (2001) administered the BDI-II to substance abuse patients. Using CFA, these researchers tested models with one, two, or three factors. Results indicated that a three-factor solution (C, A, and S) fits the data best and comprised of nine, four and eight items respectively. Johnson et al. (2006) also observed good fit for this model in a population of intravenous drug users (n = 598).

Bos et al. (2008) examined the two and three-factor structures in pregnancy and postpartum samples. The study revealed that the BDI-II 3-factor solution might be more suitable to measure depressive symptoms in pregnancy (C-A, Somatic-Anxiety and Fatigue) and postpartum (C-A, Somatic-Anxiety and Guilt). In postpartum, a three-factor solution has also been reported using a sample of recently new mothers in Kedah (Mahmud, Awang, Herman, & Mohamed, 2004). It comprised of an A factor that included items such as agitation, crying and irritability, a S factor with items such as tiredness and loss of energy and a C factor composed by items such as punishment feelings and worthlessness. Seignourel, Green and Schmitz (2007) also found a three-factor structure in treatment-seeking substance users (C, A and S). Wu and Chang (2008) too identified a three-factor solution (NA, PD, and S) in Taiwanese college students.

Not only have factor structure findings differed, but items representing the factors and the weight of item loadings within factors have also varied across studies. Several methodological studies have explored the potential for a more complex factor structure of the BDI-II using CFA. Findings from these studies generally indicate that second-order (Al-Turkait & Ohaeri, 2012; Buckley et al., 2001; Byrne et al., 2007; Grothe et al., 2005) and general-factor model structures (Thombs, Ziegelstein, Beck, & Pilote, 2008; Ward, 2006) are better fits for the BDI-II than simple two- or three-factor structures. Ward's (2006) model assumes a general (G) factor underlying the BDI-II, as well as a C (8 items) and a S factor (5 items) that are all orthogonal. Byrne and Baron (1993) tested and cross-validated the BDI-II using data from three independent samples of Canadian adolescents. The results of their study revealed the data to be most appropriately represented by a four-factor model that comprised one higher order factor of general Depression and three lower order factors that represented NA, PD, and S. Validity of this model of BDI structure has subsequently been tested for Swedish (Byrne et al., 1995) and Bulgarian (Byrne et al., 1998) adolescents and its invariance tested across gender (Byrne, Baron, & Balev, 1996; Byrne, Baron, & Campbell, 1993, 1994; Byrne, Baron, Larsson,& Melin, 1996), across the three cultural groups (see Byrne & Campbell, 1999), and across Canadian English and French cultural groups (Byrne & Baron, 1994).

2.2.3 Cross-cultural differences in the BDI-II

2.2.3.1 Cultural variations in depression

Cross-cultural research has long suggested that mental health and illness are contextually based and culturally embedded (e.g., Kleinman, 1986). According to the socio-somatic formulation in medical anthropology, "a person's context ... influences the severity and type of symptoms experienced" (James & Prilleltensky, 2002, p. 1134) and moreover, cultural categories may influence which symptoms are culturally acceptable. Cross-cultural findings on depression have also widely recognized variations in depressive symptomatology. Some cultural groups (like Nigerians and South Africans) are less likely to report extreme feelings of worthlessness. While others, like the Chinese, are more likely to report somatic complaints (Kleinman, 1988). Like other psychiatric disorders (e.g., schizophrenia), the prevalence of depression also differs from culture to culture (Marsella, 1980).

Researchers argue that cultures vary in terms of their differentiation and communication of emotional terminology, and hence in how they experience and express depression (Leff, 1977). The translation, interpretation, and meaningfulness of particular terms within various cultures are always at risk. For example, Manson (1995) noted that while the terms ‘guilt’ and ‘shame’ are typically perceived as overlapping constructs for respondents in Western cultures, they can be conceptualized quite differently by non-Western cultures (e.g., South Africa). Moreover, he pointed out that certain terms (like ‘depressed’), are totally absent from the languages of some cultures.

In arguing for a culturally relative description of depression, Kleinman writes that:

> “Depression experienced entirely as low back pain and depression experienced entirely as guilt-ridden existential despair are such substantially different forms of illness behavior with distinctive symptoms, patterns of help seeking, and treatment responses that although the disease in each instance may be the same, the illness, not the disease, becomes the determinative factor. And one might well ask, is the disease even the same?” (Kleinman, 1988, p. 25).

Although researchers generally accept the idea that depressive disease is universal, they argue that the expression and course of the illness are culturally determined (Matsumoto, 1997; Mossakowski, 2008, 2006). Marsella (1980, 1979) states, in an argument for a culturally relative perspective of depression, that depression takes a primarily affective form in individualistic cultures. In these cultures, feelings of loneliness and isolation would dominate the symptom presentation. On the other hand, somatic symptoms such as headaches would be dominant in communal cultures.

Researchers have also suggested that depressive symptom patterns will differ across cultures due to cultural variations in sources of stress as well as in resources for coping with the stress (Marsella, 1979; Mezzich & Caracci, 2008). While Beck, Weissman, Lester and Trexler (1974) described blacks and whites as having comparable mean BDI scores, while Schwab, Bialow, Brown and Holzer (1967) and Cavanaugh (1983) reported that blacks had 1-point higher mean BDI scores than whites. Nielsen and Williams (1980) further found that black females had higher mean BDI scores than males or white females, and Oliver and Simmons's (1985) indicated that non-whites scored significantly higher on the BDI-II than whites.

2.2.3.2 **Factor structure of the BDI-II across cultures**

The factorial validity of the BDI-II is still contentious, and there is no consistent assignment of items to the factors (Al-Turkait & Ohaeri, 2010; Shafer, 2006; Wu, 2010). This controversy is evident in the few reports on the factor analysis of the BDI-II from the Middle East. While one Iranian report on students supported the two-factor model (Ghassemzadeh, Mojtabai, & Karamghadiri, 2005), another Iranian study reported a five-factor solution (Ritcher et al., 1998). One study from the Arabian Gulf state of Bahrain (Al-Musawi, 2001) found three factors ("CA", "overt emotional upset", and "SV") which were much similar to the original three factors (except that the Bahraini BDI-II items: 4, 8, 10-13, 17 constituted the "overt emotional upset" domain). Al-Turkait and Ohaeri (2010) with Arab college students also supported the two-factor model. A two-factor model termed CA and S from Beck et al. (1996) has similarly been validated with Taiwan and Chinese samples (Lu, Che, Chang, & Shen, 2002; Wu & Chang, 2008). Al-Turkait and Ohaeri (2012) again found a four-factor solution (instead of the two-factor & three-factor solutions) using principal axis factoring with oblique rotation with Arab college students.

A large sample of Canadian students supported the two-factor solution similar to that from Beck's outpatient sample (with BDI-II items 1-3, 5-9 and 13-14 loading on the CA factor; while items 4, 10-12 and 15-21 loading on the SV factor) (Dozois et al., 1998). Byrne and Baron (1993) also tested and cross-validated the BDI using data from three independent samples of Canadian adolescents. The results of their study revealed the data to be most appropriately represented by a four factor model that comprised of one higher order factor of general "Depression" and three lower order factors that represented NA, PD and S. The validity of this model of BDI structure has subsequently been tested by the same researchers for Swedish (Byrne et al., 1995) and Bulgarian (Byrne et al., 1998) adolescents and its invariance tested across gender (Byrne et al., 1996; Byrne et al., 1993, 1994; Byrne, Baron, Larsson, & Melin, 1996), across the three cultural groups noted previously (Byrne & Campbell, 1999), and across Canadian English and French cultural groups (Byrne & Baron, 1994).

Byrne, Steward and Lee (2004) also found support for the four factor model with Hong-Kong adolescents. Byrne, Steward, Kennard and Lee (2007) tested measurement equivalence of the BDI-II across Hong Kong and American adolescents and found evidence of measurement equivalence of the BDI-II factorial structure across the two cultures. They also established three first-order factors of NA, PD and S and one second-order factor of general depression. Results from Byrne and Stewart (2006) also suggested that the BDI-II demonstrated factorial invariance across Hong Kong and American high school students

Two studies have compared the English and Spanish versions of the BDI measure and analyzed the BDI items for bias between Spanish and English speaking patients to determine the measurement equivalence (Azocar, Arean, Miranda, & Munoz, 2001; Penley et al., 2003). Their results supported measurement equivalence of the BDI model, but one study (Azocar et al., 2001) indicated that compared to non-Hispanic Whites, Hispanics are more likely to endorse items associated with tearfulness and punishment, and less likely to endorse the item reflecting inability to work. Lin (2012) confirmed a two-factor BDI model that emerged consistently and had a good model fit among three racial/ethnic groups (across non-Hispanic Blacks, non-Hispanic Whites, and Hispanic cardiac patients) of post- myocardial infarction patients. Similar results were found by de Jonge et al. (2006), Dunkel et al. (2002), Miles et al. (2001) and Morley et al. (2002). Another study evaluated the psychometric properties of the BDI in a sample of low-income non-Hispanic Blacks and supported MI of the two-factor (i.e., C and S) BDI model (Grothe et al., 2005). Their findings also showed that non-Hispanic Blacks were less likely to endorse the item reflecting suicidal thoughts than non-Hispanic Whites. The BDI-II has been translated into many languages, including Japanese (Kojima et al., 2002), Arabic (Al-Musawi, 2001), and Spanish (Penley et al., 2003). In factor analyses of these translated BDI-II versions, the respective researchers reported two-factor solutions similar to those reported for English versions.

Similarly, there is some evidence of mean differences on the BDI-II across groups defined by race or ethnicity (e.g., Hambrick et al., 2010; Walker & Bishop, 2005). However, Hambrick et al. (2010) found little evidence for differential item functioning on BDI-II items between White and African American undergraduate students using item response theory (IRT). However, Carmody (2005) investigated item bias in the endorsement of symptoms on the BDI-II among White, Asian, and Latino Americans. In his study, White American students scored higher on three items (BDI-II items 11, 14, and 17) than did Hispanic and Asian American students.

Hooper, Qu, Crusto and Huffman (2012) using IRT and CFA, examined scalar equivalence in responses derived from the BDI-II among 1229 college students in the United States. Results from differential item functioning analyses indicated that the items endorsed by Black American and White American college students were slightly different. Twenty- three percent of the items on the BDI-II functioned differently based on at least one comparison method (i.e., CFA or IRT). More specifically, for these race-related comparisons, symptom endorsement varied on five BDI-II items: items 7, 8, 14, 15, and 21. Therefore, five items functioned differently, and 16 of the items functioned similarly in these racial group comparisons.

Whisman et al. (2013) found evidence for MI in the context of the hierarchical four-factor structure of the BDI-II between Whites and racial (Blacks, Asians) or ethnic (Latinos) minority groups. Factor structure studies of the BDI-II with Puerto Rican participants found a four factor structure (Bernal, Bonilla, & Santiago, 1995; Bonilla, Bernal, Santos, & Santos, 2004; Lugo, 1999; Rodríguez, Joglar, & Dávila, 2005; Rosado, 1995). Another study by Rodríguez-gómez, Dávila-martínez and Collazo-rodríguez (2006) also yielded four factors with Puerto Rican elderly participants using principal component analysis as the extraction method and varimax rotation. The first factor that emerged was named "Somatic," it because included seven items regarding loss of interest, loss of energy, fatigue, and concentration difficulties. The second factor included six items related to failure, guilt, and feelings of punishment and therefore was identified as "Cognitive-behavioral". A third factor was named "Biological," that included four items regarding agitation, and problems with sleeping and appetite. The last factor, "Negative attitudes," included four items related to suicidal ideation and pessimism.

Vanheule et al. (2008) compared Beck's two factor model's fit to a one-factor model, five alternative two-factor models, and four alternative three-factor models, using CFA on data from Dutch-speaking Belgium samples of mental health outpatients and a nonclinical group. The models with a better fit were those formulated by Viljoen et al. (2003) (SA-C), Osman et al. (1997) (Negative attitude, Performance difficulty and Somatic elements), Ward (2006) (second-order and general-factor model), and Buckley et al. (2001) (CA-S). Campbell, Roberti, Maynard and Emmanuel (2009) further explored the factorial structure of the BDI-II in Anglophone Caribbean university students. Factorial estimates and goodness-of-fit indices suggested adequate fit for two-factor models. Mukhtar and Tian (2008) also reported two-factor structures (CA and SV) in a Malay sample consisting of students, general community, general medical patients, and patients with major depressive disorders. Specifically, all items loading on factor one were consistent with Beck's model, except that item 21 did not load into any of the factors in the exploratory analysis. However, item 19 in this study loaded on the CA factor whereas Beck et al. reported that the item loaded on the Somatic/Performance factor in their study.

Although most of the studies have explored the BDI–II's factor structure across cultural groups, the majority of participants were Caucasian and middle-class, with few studies in non-Western cultures (Buckley et al., 2001; Gary & Yarandi, 2004; Grothe et al., 2005). Therefore, little is known about its psychometric properties in non-Western cultures. Because cultural differences in the expression of depressive symptoms have been found and that low SES individuals are disproportionately affected by depression (Bracken & Reintjies, 2010; Hankerson, Fenton, Geier, Keyes, Weissman, & Hasin, 2011), it cannot be assumed that the validity of the BDI–II established primarily in Western cultures remain accurate for Africans. Specifically, it has been demonstrated that Africans evidence more somatic symptoms, in particular, sleep disturbance, and they articulate fewer typical depressive symptoms than depressed Western counterparts (Hankerson et al., 2011).

2.2.4 Gender differences on the BDI-II

2.2.4.1 Gender-related variances in depression

A universally held assertion, albeit tentative, is that depression is more prevalent in females than males (Beck et al., 1996; Hankin, 2002; Kessler et al., 1994; Nolen-Hoeksema, 1990). Findings from epidemiological studies offer some support for gender-related differences in depression and depressive symptoms, and suggest that this difference emerges during adolescence (see Hankin, 2002; Hankin & Abramson, 2001; Nolen-Hoeksema, 1990; Nolen-Hoeksema & Girgus, 1994; Rao & Chen, 2009). This has also been noted in the consistent evidence of substantial non-normality in the distribution of BDI item scores for nonclinical adolescents (Byrne et al., 1995; Koenig et al., 1994; Roberts et al., 1991). It is notable that the degrees of skewness and kurtosis have also varied substantially both across gender (Byrne et al., 1996; Koenig et al., 1994; Roberts et al., 1991) and across culture (Byrne et al., 1996).

However, some studies with university students have indicated that the relation between gender and depression and/or depressive symptoms is inconsistent, equivocal and unclear (Gladstone & Koenig, 1994; Nolen-Hoeksema & Girgus, 1994; Silverstein, 1999; Steer & Clark, 1997). Moreover, these gender-related variances in depression are not only unclear during emerging adulthood but also across the whole lifespan (Eaton et al., 2011; Hooper, 2010; Rao & Chen, 2009). Furthermore, the generally held claim that females have greater levels of depressive symptoms and prevalence rates of major depressive disorder is not consistently found in the empirical literature. For instance, Steer and Clark (1997) found that male college students in their study reported levels of depressive symptoms comparable to those reported by their female counterparts. In another study, Silverstein (1999) found that there are no gender differences in pure depression (i.e., without anxiety and somatic symptoms). This conclusion is consistent with findings of other empirical studies (see Gladstone & Koenig, 1994; Nolen-Hoeksema & Girgus, 1994).

2.2.4.2 **MI of the BDI-II across gender**

MI refers to the equivalence of the factor structure, item loadings onto the same factor (i.e., configural invariance), similar item thresholds (i.e., scalar invariance), and item loadings (i.e., metric invariance) in the measurement model across groups. There is a significant dearth of empirical studies on the MI of the BDI or BDI-II across gender groups (Byrne, Baron, & Campbell, 1993, Osman et al., 2004; Santor, Ramsay, & Zuroff, 1994; Wu, 2009), despite most studies having reported that females experience depressive symptoms to a greater degree and with more frequency than do males (Beck et al., 1996; Carmody, 2005; Osman et al., 1997). Boughton and Street

(2007) reviewed several studies that have found gender differences to be invariant in populations specifically of college students. Previous CFA findings show that, although a second-order factor structure best described the data for both males and females, there was some evidence of differential item functioning across gender (Byrne, Baron, & Balev, 1996).

Regarding the original inventory, that is, BDI, Byrne et al. (1993) identified one differential loading pattern (Item 20) and two nonequivalent loadings (Items 8 and 10) across adolescent males and females. Using a nonparametric item response model, Santor et al. (1994) found small but significant amounts of bias (i.e., Items 6, 10, and 14) in the BDI. Regarding the BDI-II, Osman et al. (2004) reported three noninvariance items (i.e., Items 7, 8, and 18) in boys and girls. In Wu's (2009) study, eight biased items (i.e., Items 3, 4, 5, 7, 9, 10, 16, and 17) were identified in the BDI-II when comparing adolescent males and females, with significant effects on latent mean differences. The difference in the latent means of the two genders in terms of the Cognitive-Affective domain was significant when noninvariant items were not removed, but this was not the case when these items were removed. Further factorial invariance was found for the BDI on comparisons across gender in non-clinical adolescents from Canada (Byrne et al., 1993, 1994), Sweden (Byrne, Baron, Larsson, & Melin, 1996), and Bulgaria (Byrne et al., 1996).

Wu (2010), in a sample of senior high school students, examined item bias within a two-factor model (CA and S factors). Eight noninvariant items were identified, including one nonuniform DIF item (Item 5), and seven uniform DIF items (Item 3, Item 4, Item 7, Item 9, Item 10, Item 16, and Item 17). Comparing inclusion and exclusion of noninvariant items, the results showed that the effect of noninvariance on observed mean differences was inconsequential, but it had a substantial effect on latent mean differences. Wu (2010) further explored the MI of the BDI-II in a Chinese college sample. Using a three-factor model (NA, PD, and S) as the baseline model, partial scalar invariance was supported along with five noninvariant intercepts. Males systematically reported higher scores than did females on Item 2, Item 3, and Item 4, whereas females systematically endorsed higher response scores than did males on Item 7 and Item 10. Additionally, Wu (2010) reported that the effect of noninvariance on latent mean differences was insignificant because the latent mean differences estimated separately from the model with full constraints imposed and from the model with removing five noninvariant constraints were alike.

Wu and Huang (2012) examined gender-related invariance with a Taiwanese adolescent sample within a three factor model (Negative attitude, Performance difficulty, and Somatic elements). MI was established at the level of configural, metric and partial scalar invariance. Partial scalar invariance was achieved by removing seven constraints of intercepts (Item 2, Item 3, Item 7, Item 9, Item 10, Item12, and Item 19). The intercepts of Items 2, 3 and 19 for boys were higher than those of girls, whereas the reverse held for Items 7, 9, 10 and 12. The findings of seven noninvariant intercepts implied that there was differential ARS bias (see Cheung & Rensvold, 2000) for the BDI-II across gender groups. That is, boys systematically endorse higher item responses in Items 2, 3, and 19 than girls, whereas girls systematically have higher scores in Items 7, 9, 10, and 12.

Osman et al. (1997) conducted analyses of gender differences on total scores in addition to the factor scales. In their study, a three-factor model (i.e., NA, PD, and S) was best fitted to the data of the BDI-II. Their subsequent analyses showed that females reported higher scores in terms of their overall depression (i.e., total BDI-II scores) as well as on NA and PD factors. Even so, no significant gender differences were observed for the Somatic element factor. Whisman et al. (2013) found evidence for MI in the context of the hierarchical four-factor structure of the BDI-II between women and men in college students. However, Campbell et al. (2009) in their study of the BDI-II with Caribbean students revealed no differences between males and females on the BDI-II factor scores. Hooper et al.'s (2012) study of the BDI-II MI among 1229 college students in the United States found that items endorsed by female and male college students were almost invariant. While, in Carmody's (2005) study, item-level scores differed based on gender; females had higher scores on BDI-II items 1, 10, 15, and 20.

2.2.5 **Correlations between depression and theoretically linked external criteria**

According to theory and empirical research, scores on depression should correlate with risk factors and environmental concomitants and precipitants of depression (e.g., stressful life events). Cognitive theories of depression (Beck, Rush, Shaw, & Emery, 1979) suggest that core beliefs, such as the inclination to interpret events to support negative predictions (cognitive distortions) and to attribute negative events to stable causes (hopelessness), are central to the development of depressed mood. For this reason, it is anticipated that hopelessness and stressful life events would display strong, positive correlations with depression scores (Byrne et al., 2007).

Self-efficacy is an important protective variable for depression according to the Western literature and is expected to correlate with depression scores (Bandura, 1997). Contrary to the formerly mentioned variables (stressful life events and hopelessness), the association is generally hypothesized to be negative. Self-efficacy has an indirect effect on depressive symptoms but only influences behaviours that decrease the likelihood of increased environmental stress (Bandura, 1997). Moreover, cross-cultural theory on depression suggests that, in collective cultures (e.g., South Africa), beliefs that accentuate internal sense of personal worth, efficacy, and control may be less significant than in individualistic cultures (Markus & Kitayama, 1994b); therefore, they should be weakly protective against depressed mood.

Moreover, within a diathesis-stress framework, the vulnerability model suggests that negative self-evaluations (i.e., which are conceptually close to low self-esteem; Beck, Steer, Epstein, & Brown, 1990) constitute a causal risk factor of depression (Beck, 1967; Butler, Hokanson, & Flynn, 1994; Metalsky, Joiner, Hardin, & Abramson, 1993; Stewart et al., 2004; Roberts & Monroe, 1992; Whisman & Kwon, 1993). For instance, according to Beck's (1967) cognitive theory of depression, negative self-beliefs are not just a symptom of depression but a diathesis wielding causal influence in the onset and maintenance of depression. The extant research provides unequivocal and consistent evidence for this relationship. Cross-sectional and longitudinal studies suggest that low self-esteem prospectively predicts depression (Evraire & Dozois, 2011; Hammen, 2005; Joiner, 2000; Morley & Moran, 2011; Kernis et al., 1998; McPherson & Lakey, 1993; O'Brien, Bartoletti, & Leitzel, 2006; Orth, Robins, & Meier, 2009; Orth, Robins, & Roberts., 2008; Orth, Robins, Trzesniewski, Maes, & Schmitt, 2009; Roberts & Monroe, 1992).

Despite the revision of the BDI, Beck et al. (1996) reports a high correlation (i.e., $r = 0.93$) between the original BDI and the BDI-II with their clinical sample. This finding was also replicated with a student sample by Dozois et al. (1998) ($r = 0.93$). Osman et al. (1998) reported that the BDI-II correlated substantially ($r = 0.77$) with the DASS-depression scale (Lovinbond & Lovibond, 1993). Aasen (2001) further reported high and significant correlations with the SCL-90-R depression subscale and the Zung Self-Rating Depression Scale. Mukhtar and Tian's (2008) study also revealed a significant positive correlation between BDI-Malay total scores with Zung ($r = 0.80$). Al-Turkait and Ohaeri (2010) recently examined the convergent validity of the BDI-II with the Hopkins Symptom Checklist (HSCL-25) in Arab college students. All correlations with the HSCL-25 domain scores were highly significant (r mostly > 0.50, $p < 0.001$). The summed scores of the cognitive factors of the two-factor models had significantly higher correlations with the depression score of the HSCL-25 ($r = 0.66$-0.70) than with the HSCL-25 anxiety score ($r = 0.54$-0.57) ($Z = 3.9$, $p < 0.001$).

Essentially, the correlations between the BDI-II and discriminant measures are expected to be negative as opposed to correlations between the BDI-II and convergent measures. There are contradictions in the literature on the relationship between depression symptoms and anxiety symptoms. According to a cognitive model of depression and anxiety, the two disorders can be separated by their cognitive content, which is loss for depression and fear for anxiety (Beck, 1967; Beck et al, 1979). Despite this theoretical separation, there is substantial evidence for the comorbidity of the two disorders (Maser & Cloninger, 1990). Researchers have also found that it is difficult to differentiate the two disorders' symptoms in non-clinical samples (Gotlib, 1984; Joiner, 1996). The BAI (Beck & Steer, 1993a) is however often used as a discriminant measure in most BDI studies (Beck et al., 1988). Osman et al. (1997) reported a correlation of $r = 0.56$ between the BDI-II and BAI, which is similar to that by Steer and Clark (1997). Beck et al. (1996) also reported a correlation of $r = 0.47$ between the BDI-II and the Hamilton Anxiety Rating Scale (Hamilton, 1959) with an outpatient sample.

2.2.6 **Longitudinal MI of the BDI-II**

There is generally a paucity of studies that have explored the longitudinal MI of the BDI-II. Beck et al. (1996) examined the test-retest reliability of the BDI-II over one-week in a sample of psychiatric outpatients and reported correlations of $r = 0.93$. Aasen (2001) also investigated the test-retest reliability of the BDI-II in a student sample over a three-week period and found a test-retest correlation of $r = 0.77$.

Byrne et al. (2004) tested and cross-validated the BDI-II higher-order factorial structure (Beck et al., 1993, 1995, 1998) with two independent samples of Hong Kong adolescents. Consistent with past research, findings revealed this hierarchical structure to fit the data exceptionally well and to be invariant across the six-month time lag. This finding of stability of the BDI-II across time by Byrne et al. (2004) is consistent with those reported by past studies (Beck et al., 1996; Dozois et al., 1998).

2.3 **Conclusion**

This section discussed the different factor structures (e.g., two, three and hierarchical factor structures) of the BDI-II across populations, gender and time. Moreover, the correlations between depression and theoretically linked external criteria were explored. It is evident from the above-mentioned research that factor analytic studies on the BDI-II appear inconsistent and at times contradictory (i.e., factor solutions and item loadings). It is these inconsistencies in the literature that have also motivated the present investigation.

CHAPTER 3
METHODOLOGY

3. Introduction

This chapter outlines the methods adopted in this study. This will also include discussions on the data collection instruments and procedures followed in conducting the research.

3.1 Study design

A MTMM methodology is adopted for this study, encouraged by the recommendations of Campbell and Fiske (1959). Multiple methods (i.e., questionnaire and interview schedule) were used in the validation process to ensure that the variances reflected are those of the trait and not of the method. Thus, the convergence or agreement between two methods serves as cross validation/triangulation and enhances the confidence that the results are valid and not a methodological artifact. The MTMM does afford an expedient way of reporting reliability and construct validity coefficients of an instrument (Campbell & Fiske, 1959). In this study, the construct validity of the BDI-II was examined through the EFA and CFA. This was followed by the MTMM based approach to examine the relative magnitudes of the within-trait, between-method correlations, the between-trait, within-methods correlations, and the between-trait, between-method correlations for convergent and discriminant validity.

The survey was contemporaneous and took place along two time periods (i.e., over a 2-week time lag). This enabled factors of interest (i.e., cognitive, affective and somatic) to be examined for stability over time. It also examined change within individuals as well as variation between them. Repeated measures further allowed for the detection of change in individuals or their environments from one data point to the next. This helped determine internal consistency reliability related to the total scale (i.e., BDI-II), as well as to each factor or subscale, at each of two time points; stability of the lower-order factors over a 2-week time lag; invariance of the lower-order factor loadings, and the measurement error variances over time.

3.2 **Participants**

A purposive heterogeneous sample was drawn for this study from both the University of Limpopo and University of Pretoria, in South Africa. The first phase of the data comprised of 919 students. Four hundred and twenty five (50.1%) first-phase questionnaires were collected from the University of Limpopo, while 493 (53.7%) came from the University of Pretoria. Participants were undergraduate students, aged 17 to 50 years, with a mean age of 21.70 yr. (SD = 13.51). Of all the participants in the first phase, 579 (63.4%) selected the classification of 'Black', 291 (31.9%) were 'White', 26 (2.8%) were 'Asian' and 17 (1.9%) were 'Coloured' (The classifications are now commonly used as sociological constructs in South Africa). Only 304 (33.08%) of the total follow-up questionnaires distributed were returned.

The rationale for selecting this student population is that it approximates to the ones used in previous studies on the BDI-II (Beck et al.,1996; Byrne et al., 2004) and further aided in the testing of the equivalence of the determined factor structure between two cultural groups (i.e., Black and White).

3.3 **Sample size adequacy**

A minimum of 5:1, and preferably 10:1, subject-to-item ratio has been a traditional standard for EFA sample size requirements. However, recently, empirical findings using Monte Carlo simulations have advanced sample size estimation methodology for conducting EFA (Gorsuch, 1983; Kneipp et al., 2009). It was specifically established that, where communalities are moderate to high (≥ 0.40), when there is a small number of factors, and when factors are over-determined (6 or more indicators per factor), a sample size of 100 (and in some cases less) is sufficient, regardless of the subject-to-item ratio (MacCallum, Widaman, Preacher & Hong, 2001; MacCallum, Widaman, Zhang & Hong, 1999). The sample in the present study meets the preferable 10:1 subject-to-item ratio for EFA.

Comparable to the sample size estimation for EFA, traditional sample size recommendations for CFA are 500 subjects. But a compromise figure of between 100 to 200 subjects has been advanced due to the difficulty of obtaining 500 subjects (Kneipp et al., 2009). However, recent findings suggest sample size adequacy is dependent on key qualities of the CFA model, such as the number of indicators per factor, item loadings, and the number of factors included (Gagne & Hancock, 2006). The aim of sample size estimation for CFA is to achieve satisfactory model convergence with fewer iterations (Gagne & Hancock, 2006).

The present research sample meets the recommended sample size estimation for the conventional standards of CFA (n = 460 & 459) of the model based on the guidelines of Gagne and Hancock (2006).

3.4 **Instruments**

The data was collected with the following instruments: a demographic questionnaire, BDI-II, Hopkins Symptom Checklist-25 (HSCL-25), Structured Clinical Interview for DSM-IV-TR diagnosis (SCID-I/NP), Hopelessness Scale, Sherer General Self-Efficacy Scale (SGSES), Rosenberg Self-Esteem Scale (RSES), and the Perceived Stress Scale (PSS-10).

3.4.1 *Demographic questionnaire*

In the demographic information, participants were asked to provide information on their background and current family situation. All the respondents indicated their age, gender, ethnic identification and SES.

3.4.2 *BDI-II*

The BDI-II is a well-validated self-report questionnaire (largely in North America and Europe) comprising of 21 items (Beck & Steer, 1987). It measures the severity of depressive symptoms in adolescents and adults. Beck originally hypothesized that thirteen items cover cognitive and affective components of depression such as pessimism, guilt, crying, indecision, and self-accusations; eight items assess somatic and performance variables such as sleep problems, body image, work difficulties, and loss of interest in sex. The examinee receives a score of 0 to 3 for each item; total raw score is the sum of the endorsements for the 21 items; the highest possible score is 63.

In a meta-analysis of the BDI research studies, the internal consistency of the scale (mainly coefficient alpha) ranged from $\alpha = 0.73$ to 0.95, with an average internal consistency of $\alpha = 0.86$ in nine psychiatric populations (Beck, Steer & Garbin, 1988). The BDI-II generally possesses adequate internal consistency, with a coefficient alpha of 0.84 to 0.92 in previous studies (Beck, Steer, & Brown, 1996) and 0.84 in the present study. Test-retest reliability of the BDI is modest, with a range of $r = 0.60$ to 0.83 in non-psychiatric samples and $r = 0.48$ to 0.86 in psychiatric samples. Its correlations with clinical ratings and scales of depression from measures such as the MMPI are typically in the range of $r = 0.60$ to 0.76 (Conoley, 1992).

3.4.3 *HSCL-25*

The HSCL-25 (Mollica, Wyshak, de Marneffe, Khuon, & Lavelle, 1987) is a widely used screening self-report measure that includes symptoms of anxiety (10 items) and depression (15 items) derived from the 90-item Symptom Checklist (SCL-90). It consists of two subscales: the anxiety (HSCL-10) and depression (HSCL-15) subscales. The measure is scored on a severity scale from '1' (not at all) to '4' (extremely), and has demonstrated its usefulness as a screening tool in various cross-cultural settings (Kaaya, Fawzi, Mbwambo, Lee, Msamanga, & Fawzi, 2002). Respondents scoring higher than a mean of 1.75 on the HSCL-25 (full scale), the HSCL-10 (anxiety), or the HSCL-15 (depression) are classified as having significant emotional distress (Mollica et al., 1987).

The HSCL-25 is comparable with other assessment instruments such as the CES-D in detecting the presence of a psychiatric disorder (Radloff, 1977). The full scale HSCL-25 has a demonstrated internal consistency of 0.95 (Coyne, 1994) and correlates highly with the standard 58-item version of the scale (HSCL-58; Derogatis, Lipman, Rickels, Uhlenuth, & Covi, 1974). Previous studies have reported internal reliabilities of 0.84 (HSCL-10) and 0.91 (HSCL-15) for the subscales (Kaaya et al., 2002; Syed, Zachrisson, Dalgard, Dalen & Ahlberg, 2008). The HSCL-25 displayed an adequate Cronbach's alpha of 0.96 in South African studies (Halvorsen & Kagee, 2010; Kagee, 2005; Kagee & Martin, 2010). The full scale and the two subscales of HSCL-10 and HSCL-15 also displayed high internal consistencies of α = 0.89, 0.81 and 0.84 respectively, in the present study.

3.4.4 *SCID-I/NP*

The SCID-I/NP (First, Spitzer, Gibbon & Williams, 2002) is a semi-structured interview schedule with multiple modules that reflect DSM-IV-TR defined disorders. The interview reflects the clinical diagnostic process that is employed by trained clinicians. The SCID-I/NP has demonstrated good validity and reliability among English speaking populations (Kaaya et al., 2002). Symptoms are rated in a format that allows for rephrasing and asking additional clarifying questions, an aspect that affords helpful flexibility when its use is transferred to a cultural context that differs from its source. The SCID-I/NP modules are used to characterize common mental disorders at primary care levels and include those for major depressive disorder, generalized anxiety disorder, mixed anxiety depressive disorder, and somatization disorder. The modules elicit reports on lifetime prevalence of DSM-IV-TR disorders, as well as 1-year and 1-month prevalence (Kayaa et al., 2002). The inter-rater and internal reliabilities of the SCID-I vary from adequate to excellent (Martin, Pollock, Bukstein, & Lynch, 1999; Zanarini & Frankenburg, 2001; Zanarini et al., 2000). The SCID-I/NP for DSM-IV-TR Axis I disorders was used for this study (e.g., Module A: Depression-both minor and major).

3.4.5 *SGSES*

The SGSES (Sherer, Maddux, Mercandante, Prentice-Dunn, Jacobs, & Rogers, 1982) is a Likert format 17-item scale. The response format is a 5-point scale (1 = strongly disagree, 5 = strongly agree). The sum of item scores reflects general self-efficacy. The higher the total score is, the more self-efficacious the respondent. Sherer et al. (1982) developed the SGSES scale to measure "a general set of expectations that the individual carries into new situations" (p. 664). The SGSES has been the most widely used of the self-efficacy measures (Chen, Gully & Eden, 2001). The SGSES was primarily developed for clinical and personality research.

Reviewing various studies, Chen et al. (2001) found the internal consistency reliabilities of SGSES to be moderate to high (α = 0.76 to 0.89). Likewise, the scale's reliability was also moderate in the present study (α = 0.53). In two of their studies using samples of university students and managers, Chen et al. (2001) reported a high internal consistency reliability for the SGSES (α = 0.88 to 0.91, respectively). With regard to temporal stability of the SGSES, Chen, Gully, Whiteman and Kilcullen (2000) obtained a low test-retest reliability estimate (r = 0.23) across only 3 weeks. However, Chen et al. (2001) found high test-retest reliability (r = 0.74 and 0.90). Research results show that SGSES negatively correlates with negative affect, anxiety, depression, anger, and physical symptoms (e.g., Leganger, Kraft, & Røysamb, 2000; Luszczynska, Gutiérrez-Donã, & Schwarzer, 2005).

3.4.6 *RSES*

The RSES (Rosenberg, 1965) is a 10-item Guttman scale that refers to self-respect and self-acceptance rated on a 4-point Likert-type scale, ranging from 1 (totally disagree) to 4 (totally agree). Items 1, 3, 4, 7, and 10 are positively worded, while items 2, 5, 6, 8, and 9 are negatively worded. The RSES scale is the most widely used measure of global self-esteem (Hagborg, 1993; Shahani, Dipboye, & Phillips, 1990). It was used in 25% of the published studies reviewed by Blascovich and Tomaka (1991). The RSES had a high internal reliability in previous studies (α = 0.92) (Hagborg, 1993) and also a moderate internal consistency at 0.73 in the present study.

3.4.7 *PSS-10*

The PSS-10 (Cohen & Williamson, 1988) is a self-report instrument designed to assess the degree to which situations and circumstances in one's life are appraised as stressful. It was designed to tap how unpredictable, uncontrollable and overwhelming respondents find their lives. The PSS-10 requires participants to respond to a series of questions using a 5-point Likert scale (never; almost never; sometimes; fairly often; very often). Higher scores reflect greater stress levels. The PSS-10 has high coefficient alpha reliabilities, generally 0.75 and above (Baldwin, Harris, & Chambliss, 1997). Campbell et al. (2009) further reported a high Cronbach's alpha for the PSS-10 at 0.79. However, the scale displayed a moderate internal consistency in the present study (α = 0.55).

3.4.8 *BHS*

The BHS (Beck et al., 1974) was designed to measure three major aspects of hopelessness: feelings about the future, loss of motivation, and expectations. The test is designed for adults, aged 17-80, and consists of a list of 20 statements. The person is asked to decide about each sentence whether it describes his/her attitude for the last week, including the day in question. If the statement is false for him, he should write "false" next to it. If the statement is true for him, he should label it "true". There are seven reversed items: 1, 5, 6, 8, 13, 15 and 19. Scores of 4—8 indicate mild hopelessness, 9—14 moderate and 15—20 severe hopelessness (Forintos, Sallai, & Rózsa, 2010). The BHS has high coefficient alpha reliabilities of 0.82 to 0.93 (Beck et al., 1974; Forintos et al., 2010). The instrument displayed a moderate internal reliability in the present study (α = 0.69).

3.5 **Procedure**

The sample consisted of a heterogeneous group of non-clinical student respondents. Students were recruited from undergraduate classes at the University of Limpopo and the University of Pretoria, after permission for access was granted by both institutions (see appendix b). The purpose of the research was first explained to the participants and they were further made aware that participation in the study was voluntary. Informed consent (appendix b) was then obtained from those students who wished to participate in the study, after which the instrument was administered. Instructions were further given as to how the questionnaire was to be completed. Lastly, confidentiality and anonymity was assured. The measures were administered to all participants on two occasions with a time lag of 2 weeks (Time 1: N = 919 & Time 2: N = 304). The questionnaire was administered in a group set-up in both occasions. The researcher and trained assistants were available at all the data collection halls, to off any further clarifications and answer any new questions that arose during the process.

The researcher followed ethical guidelines as specified by the Research and Ethics Committee, University of Limpopo. This included requesting ethical approval and undergoing an ethics review process before engaging participants, to ensure that procedures were fair and unbiased and not harmful to all involved. As already stated, participants completed an informed consent form before taking part in the study as per the recommendations of the Health Professions Council of South Africa (Allan, 2001). Furthermore, the researcher debriefed the participants and provided information regarding existing treatment resources (i.e., university student counselling centres) to any participant who felt affected by the process.

3.6 Conclusion

This section provided an exposition of the research design, procedures the researcher followed and participants used in the study.

CHAPTER 4
RESULTS

4. Introduction

This chapter presents analytic processes, methods, results and interpretation of the data for the current study.

4.1. Data analyses plan

The analyses were conducted in several stages using the SPSS 22.0 and EQS 6.1 (Bentler, 2004) programs. Data screening and "cleaning" (i.e., vetting data for capturing mistakes and the handling of missing data) was performed, and descriptive statistics (i.e., mean, standard deviation, percentages, skewness and kurtosis) were computed and presented to provide an overall picture of the data obtained. Due to the large sample and together with our intent to test aspects of BDI-II structure from varying perspectives that involved several analytic procedures, it was ideal to work with data that were complete both within and across time. Standard mean imputation for missing values with the SPSS was conducted in the current study. The decision to implement mean replacement of missing data was based on the following criteria: (a) although the data were not missing completely at random (NMAR) (see Little & Rubin, 1987), as evidenced by significant results derived from the Generalized Least Squares test of homogeneity of covariance matrices representing complete and incomplete data, $\chi^2(55{,}794, N = 909) = 40$, $p = 0.05$, and (b) the amount of missing data across two time points was less than 5%.

Following data "cleaning", correlation analysis, factor analysis and SEM were also computed to test the hypotheses. This was followed by the internal consistency/stability of the instruments which were conducted on the derived factors. The test of MCFA for MI was also performed (see Bollen, 1989; Meredith, 1993). The analyses of each stage are now fully described:

Stage 1.

The data were first randomly split into two independent groups and then tested using one of two factor analytic approaches. For Group 1 (n = 460), we applied EFA using principal component analysis with varimax rotation. Rotation method was chosen in accordance with Tabachnick and Fiddell's (2007) recommendation that:

> "Perhaps the best way to decide between orthogonal and oblique rotation is to request oblique rotation [e.g., direct oblimin or promax from SPSS] with the desired number of factors [see Brown, 2009b] and look at the correlations among factors…if factor correlations are not driven by the data, the solution remains nearly orthogonal. Look at the factor correlation matrix for correlations around 0.32 and above. If correlations exceed 0.32, then there is 10% (or more) overlap in variance among factors, enough variance to warrant oblique rotation unless there are compelling reasons for orthogonal rotation" (Tabachnick & Fiddell, 2007, p. 646).

However, in our analysis with direct oblimin, the correlation coefficients among the factors were not substantial, that is, two of the three coefficients were < 0.32, and hence we proceeded with varimax rotation. Factor selection was guided by a set of standard criteria, including: 1) the Kaiser-Guttman rule (i.e., factors with eigenvalues of ≥1.0), 2) the scree plot, 3) cumulative and unique percent of explained variance, and 4) prior EFA findings. Using the three-factor lower-order structure turned out to be most appropriate for Group 1, and heeding the recommendations of Byrne and Baron (1993) that the BDI structure is most appropriately represented as a hierarchical structure, we tested next for the validity of the lower-order factor structure for Group 1 (i.e., the same group) (n = 460) using CFA within the framework of SEM.

More specifically, the model specified the three lower-order factors determined in the EFA, albeit with the addition of a single higher-order factor of Depression. Several criteria were used in determining the goodness of fit to the data for this hypothesized structure; these included the Yuan-Bentler scaled Chi-square test (Y-B χ^2), Satorra-Bentler scaled Chi-square test (S-B χ^2), Comparative Fit Index (CFI), Bentler-Bonett Non-Normed Fit Index (B-B NNFI), the Standard Root Mean Squared Residual (SRMR), and the Root Mean Square Error of Approximation (RMSEA), along with its related 90% confidence interval (the evaluative criteria related to each is detailed later in the results section). Finally, using CFA again, the best-fitting model for Group 1 was cross-validated with Group 2 data (n = 459). All subsequent testing of the data were based on the full sample (N = 919).

Stage 2. As a final validity check of our postulated structure of the BDI–II, we assessed if there will be any gender and race differences on the BDI-II, to determine the generalizability of psychometric data across male and female, and white and black South African subsamples. The levels of MI (i.e., configural, metric and scalar) were examined with a MCFA, which allows for testing a priori theory of the test structure across groups. This approach allows for the comparison of specific features of the factor model from one group to another. When the features are found to be equivalent across groups, MI, specifically factorial invariance, can be inferred.

Stage 3. Stability of the lower-order factor loadings was assessed using analysis of covariance structures to test for their invariance across the two time points.

Stage 4. Based on the final best-fitting model from Stage 1, we determined the internal consistency reliability, at two time points, for the total scale and for each of the three lower order factors.

Stage 5. Correlations between depression and theoretically linked external criteria were examined. Based on theory and empirical research, scores on depression were expected to correlate with risk factors as well as with environmental precipitants.

Alternative models previously established elsewhere were also examined to form comparisons with our hypothesized model:

Stage 6. Byrne et al.'s (2004) three-factor lower-order model, comprising of Negative attitude, Performance difficulty and Somatic complaints factors was also tested for our sample, together with gender, race and longitudinal MI.

Stage 7. Furthermore, Beck et al.'s (1996) two-factor model, comprising of a Cognitive-affective factor and a Somatic factor was lastly tested for our sample.

4.2 **Presentation of results**

4.2.1 *Descriptive data*

The percentages of symptom ranges for the student sample on the total BDI-II were: 65% of the students scored at a minimal symptom range, 20% on mild range, 11% on moderate symptom range and 4% on severe symptom range. The mean score (11.45; SD = 7.74) of the sample on the total BDI-II is outside the symptomatic range of the BDI-II suggested by Beck et al. (1996). The frequency distribution of the BDI-II total scores approximated a non-normal distribution, Kolmogorov-Smirnov $z = 0.09$, $p < 0.001$. In addition, skewness and kurtosis were also calculated to further determine the normality of the data (see table 1). The total BDI-II skewness was < 3 for both time 1 and 2, whereas the kurtosis was < 3 for time 1 but > 3 for time 2. Item skewness and kurtosis were variable across time 1 and time 2. The skewness of nineteen items was < 3 for both time 1 and time 2 while > 3 for the remaining two items. The kurtosis for the sixteen items was < 3 for both time 1 and 2 while > 3 for the remaining five items. This also demonstrated that the distribution of the data is non-normal for both times and that the interpretations of all analytic work in this study should be based on the robust statistics. The item-total correlations ranged from 0.26 for item 21 (loss of sexual interest) to 0.51 for item 14 (worthlessness). The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and the Bartlett test of sphericity were superb at 0.90, $p < 0.001$

(Hutcheson & Sofroniou, 1999). The determinant value for these data was 0.009, which was greater than the necessary value of 0.00001. Therefore, multicollinearity did not emerge as a problem for this data. All items in the BDI-II correlated fairly well with each other and the total score, and none of the correlation coefficients were particularly large (i.e., > 0.9, multicollinearity); therefore, there was no need to consider eliminating any items.

Table 1:

BDI-II statistics for the full sample

		M		*SD*		*Skewness*		*Kurtosis*	
	Item descriptor	*T1*	*T2*	*T1*	*T2*	*T1*	*T2*	*T1*	*T2*
1.	Sadness	0.38	0.30	0.61	0.61	1.84	2.56	4.31	7.63
2.	Pessimism	0.24	0.15	0.51	0.41	2.36	3.35	6.40	15.07
3.	Past Failure	0.32	0.22	0.62	0.52	1.89	2.44	2.95	5.74
4.	Loss of pleasure	0.56	0.42	0.73	0.61	1.16	1.25	0.78	0.95
5.	Guilty feelings	0.58	0.38	0.61	0.59	0.68	1.35	0.29	1.32
6.	Punishment feelings	0.41	0.22	0.86	0.59	2.08	3.11	3.15	9.98
7.	Self-dislike	0.40	0.30	0.68	0.62	1.51	2.13	1.23	4.15
8.	Self-criticalness	0.72	0.43	0.93	0.74	1.13	1.74	0.24	2.45
9.	Suicidal thoughts	0.12	0.10	0.37	0.36	3.63	4.42	17.13	23.11
10.	Crying	0.74	0.51	1.09	0.95	1.17	1.77	-0.15	1.78
11.	Agitation	0.55	0.43	0.79	0.68	1.62	1.82	2.31	3.68
12.	Loss of interest	0.62	0.42	0.78	0.62	1.23	1.64	1.12	3.29
13.	Indecisiveness	0.58	0.37	0.81	0.57	1.44	1.34	1.54	1.42
14.	Worthlessness	0.28	0.24	0.63	0.66	2.19	2.76	3.80	6.79
15.	Loss of energy	0.66	0.56	0.68	0.67	0.74	0.93	0.26	0.29
16.	Changes in sleeping Pattern	1.05	0.82	0.86	0.89	0.61	1.45	-0.14	4.03
17.	Irritability	0.53	0.43	0.72	0.64	1.34	1.44	1.44	1.85
18.	Changes in appetite	0.70	0.58	0.68	0.69	0.90	1.09	1.32	1.07
19.	Concentration difficulty	0.79	0.55	0.82	0.69	0.56	0.87	-0.83	-0.49
20.	Fatigue	0.76	0.54	0.73	0.63	0.84	0.98	0.76	1.06
21.	Loss of interest in sex	0.48	0.34	0.89	0.74	1.75	2.22	1.79	4.12
	BDI-II scale	11.45	8.30	7.74	8.08	0.89	1.60	0.81	3.46

Note. T1 = Time 1, T2 = Time 2; Time 1 N = 919; Time 2 N = 304:

The values stated in table 4 and 12 among others; denote what Bentler (2004) terms as “robust statistics.” That is, they have been scaled (or corrected) to account for some non-normality in the data. Given that the data comprise of responses from university students (i.e., a nonclinical population), evidence of both skewness and kurtosis (i.e., normality if skewness and kurtosis have values between –1.0 and +1.0) is indeed not unexpected and is consistent with other BDI studies of community samples in general and adolescents in particular (e.g., Byrne et al., 1995; Byrne et al., 2004; Koening et al., 1994; Roberts et al., 1991). Hence, this aspect of the data should be of little concern. However, given these distributional characteristics (see table 1), what is imperative is that analyses be based on the correct statistics, that is, they should be based on algorithms designed to take this non-normality into account. EQS 6.1 has the option of using robust statistics, rather than the regular statistics to address this problem. As a result, interpretation of all analytic work in this present study was based on the robust statistics.

4.2.2 *Stage 1 analyses: EFA and CFAs for the hypothesized model*

4.2.2.1 *EFA of the BDI-II*

EFA performed on Group 1 (n = 460) data, yielded a three-factor solution (see table 2). Eigenvalues were 3.25 (15.46 % of variance), 2.75 (13.10 % of variance) and 2.38 (11.33 % of variance) for the first, second and third factors, respectively. All items loaded significantly on three factors that could be appropriately labelled as Negative attitude, Performance difficulty, and Somatic complaints (consistent with Byrne et al. 1999, Byrne et al. 2004; Osman et al., 1998). All 21 items demonstrated acceptable factor loadings (≥ 0.30) on a given factor following rotation, with the vast majority loading at 0.40 or higher (see Kneipp et al., 2009; Pett et al., 2003).

Only one item cross-loaded on more than one factor like in Steer and Clark (1997): Item 1 (sadness) with loadings of 0.44 on Factor 1 (Negative attitude) and 0.42 on Factor 3 (Somatic complaints). All factor loadings are reported in table 2; the italicized factor loadings represent the item considered to cross-load on two factors. Thus this item was considered not specific to any domain of depression in our sample. As noted by Ward (2006), in existing factor analytic studies of the BDI-II, the so-called cognitive or somatic factors also contain affective or emotional content and maybe this phenomenon accounts for the occurrence of these cross-loadings. Otherwise, all items loaded distinctively and without cross-loadings. Heeding Costello and Osborne's (2005) recommendations on factor analysis, that after rotation and comparison of the item loading, the model with the "cleanest" factor structure–item loadings above 0.30, few or no item cross-loadings, no factors with fewer than three items–has the best fit to the data, we considered our three-factor model to have the best fit for the student data.

Table 2:
EFA three-factor solution

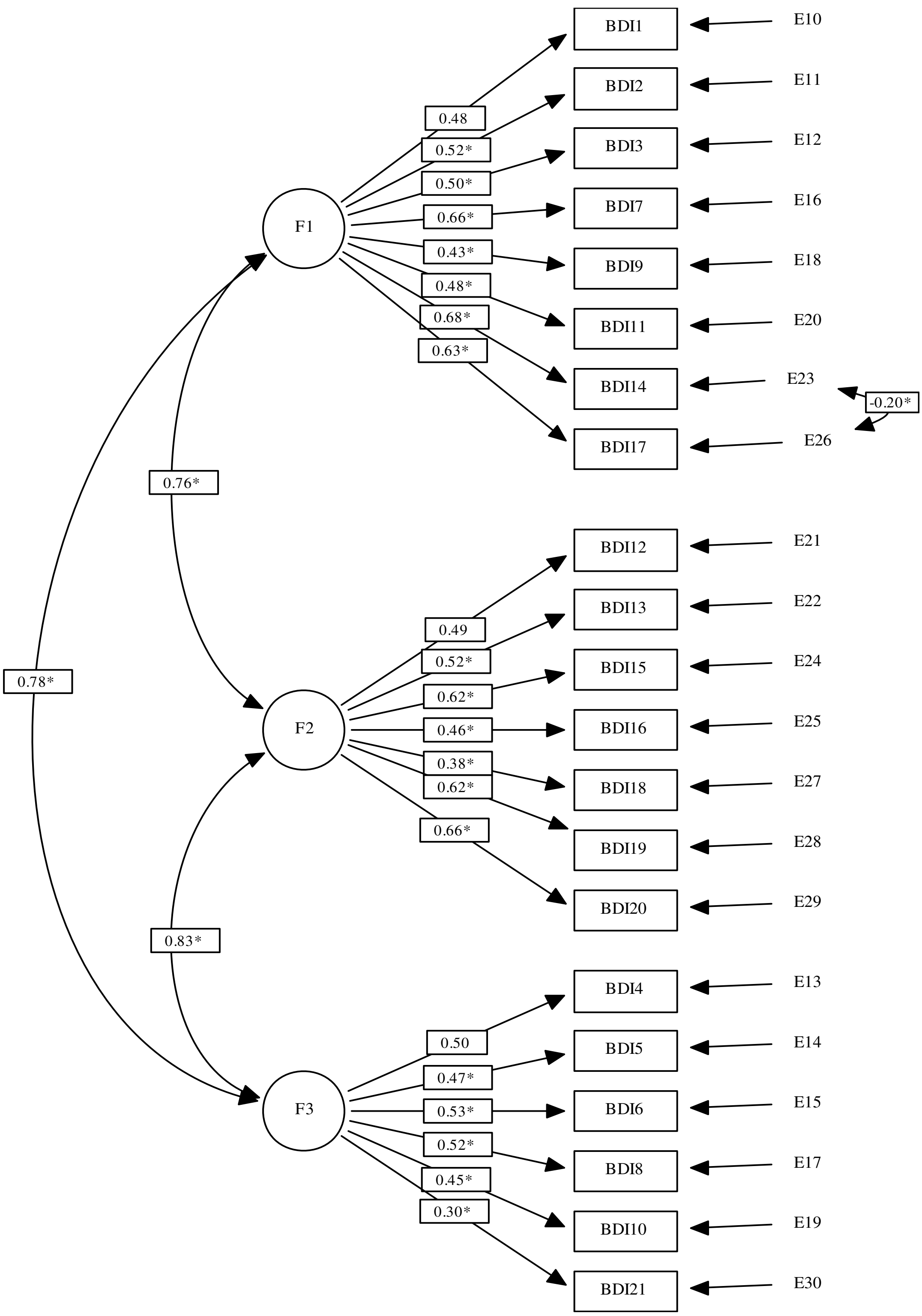
Item descriptor	*Factor 1- Negative attitude*	*Factor 2- Performance difficulty*	*Factor 3- Somatic complaints*
1. Sadness	***0.44***		***0.42***
2. Pessimism	0.73		
3. Past failure	0.57		
4. Loss of pleasure			0.48
5. Guilty feelings			0.43
6. Punishment feelings			0.49
7. Self-dislike	0.67		
8. Self-criticalness			0.32
9. Suicidal thoughts	0.50		
10. Crying			0.63
11. Agitation	0.42		
12. Loss of interest		0.43	
13. Indecisiveness		0.40	
14. Worthlessness	0.62		
15. Loss of energy		0.64	
16. Changes in sleeping pattern		0.68	
17. Irritability	0.56		
18. Changes in appetite		0.53	
19. Concentration difficulty		0.51	
20. Fatigue		0.64	
21. Loss of interest in sex			0.66

4.2.2.2 *CFA 1*

Based again on data from Group 1 (n = 460), the calibration sample, the validity of the BDI–II structure derived empirically with the EFA and schematically portrayed in figure 1 was tested. Results revealed a well-fitting model to the data. Goodness-of-fit statistics related to the test of this hypothesized model are shown in table 3.

Figure 1:

Pattern coefficients for the three-factor model



Note. The values on the path diagram are standardized regression coefficients, with arrows pointing from latent variables to the observed variables.

Table 3:

Three factor model of the BDI-II structure: Goodness-of-fit statistics

Group	*df*	*B-B NNFI*	*S-B χ^2*	**CFI*	**RMSEA*	*90% *RMSEA CI*	*SRMR*
Calibration[a]	185	0.92	275.48	0.93	0.03	0.02, 0.04	0.05
Validation[b]	390	0.93	546.89	0.93	0.03	0.02, 0.04	0.05

Note. [a]n = 460, [b]n = 459; df = degrees of freedom; B-B NNFI = Bentler-Bonett Non-Normed Fit Index; S-B χ^2= Satorra-Bentler scaled Chi-square test; *CFI = robust Comparative Fit Index; *RMSEA = robust root mean square error of approximation and its 90% confidence interval; SRMR = standardized root mean-square residual.

The SEM testing was based on Robust Maximum Likelihood (ML) estimation using the EQS 6.1 for Windows program (Bentler, 2005). To identify the most significant and meaningful model modifications, we examined the Lagrange Multiplier (LM) tests and added paths that were most likely to improve the fit of the model and which made theoretical sense. To evaluate the fit of the models we focused on different types of fit indices, including the B-B NNFI, CFI, SRMR and RMSEA. Following convention, for example, Byrne (2006) models with B-B NNFI and CFI values greater than 0.90, a SRMR less than or equal to 0.05 and a RMSEA less than or equal to 0.10 are judged as providing a reasonable fit to the data. In this study, however, Hu and Bentler's (1999) recommendation was used, that is to accept the model if CFI $\geq$ 0.95 and RMSEA < 0.06. Models with CFI between 0.90 and 0.95 were also accepted due to the demonstrated difficulty of achieving optimal cut-off values with small samples (see Sivo, Fan, Witta & Willse, 2006). As noted by Sivo et al. in their study:

> "…Just as fit indexes are affected by sample size, optimal cut-off values (for correct models only) vary considerably depending on sample size, with smaller sample sizes resulting in lower optimal cut-off values. If our interest is to retain all correct models (i.e., no Type I error) while maximizing the chances of rejecting misspecified models (i.e., minimizing Type II error), then the cut-off values for all indexes could become more stringent as sample size increases from 150 to 5,000 (see table 3). This result suggests that larger sample sizes offer more precision in identifying the correct (i.e., true) model. This finding suggests that, regardless of which index is under consideration, the cut-off values may need to become less rigorous as sample size decreases, so that we could retain all correct models while maximizing the chance of rejecting the incorrect models as rival hypotheses" (Sivo et al., 2006, pp. 284-285).

We also examined the Y-B χ^2 and S-B χ^2. For a good model fit the Chi-square statistic value should not be significant. However, since the Chi-

Square test is sensitive to large sample sizes we used the ratio of Y-B χ^2 and S-B χ^2 to model degrees of freedom as a model criterion. A ratio of 1.5 or below was considered as an indication of adequate model fit.

The first step in SEM analysis was to test for the hypothesized model (i.e., a three-factor lower-order model generated from the data using EFA). For the structural model, the significant χ^2 value and its ratio to df is less than 1.5 (S-B $\chi^2 = 275.477$, df = 185, $p < 0.0001$, $\chi^2/\mathrm{df} = 1.49$), along with goodness-of-fit indices, namely, the B-B NNFI = 0.92 and CFI = 0.93, which were slightly less than the cut-off point suggested by Hu and Bentler (1999), but nonetheless indicative of a well-fitting model (applying criteria used by Dozois et al. [1998] and Osman et al. [1998]), and an acceptable SRMR = 0.05 and RMSEA = 0.03 with a 90% CI for RMSEA of 0.02- 0.04, implied that the model fits the data well. The model for the study is presented in figure 1.

The parameter estimates of the SEM (figure 1) are shown in table 4. All path coefficient estimates have the expected signs. The magnitudes of the standardized path coefficient estimates, for the measurement components of the model, suggest that the items BDI-2, BDI-3, BDI-7, BDI-14 and BDI-17 have a stronger effect on Factor 1: “Negative attitude” than BDI-1, BDI-9, and BDI-11, and BDI-13, BDI-15, BDI-19 and BDI-20 have a stronger effect on Factor 2: “Performance difficulty” than BDI-12 and BDI-4, BDI-5, BDI-6, and BDI-8 has a stronger effect on Factor 3: “Somatic complaints” than BDI-10 and BDI-21. Moreover, the model also contains correlations between residual variances.

Table 4:

Maximum likelihood estimates of measurement and structural path coefficients for the calibration data (n = 460)

Items	*Coefficient (Robust SE)*	*Robust t – ratio**	*Standardized coefficient*
F1			
BDI-1	1.00		0.48
BDI-2	0.91 (0.17)	5.51	0.52
BDI-3	1.03 (0.22)	4.78	0.50
BDI-7	1.50 (0.22)	6.90	0.66
BDI-9	0.52 (0.13)	4.01	0.43
BDI-11	1.22 (0.18)	6.78	0.48
BDI-14	1.41 (0.20)	7.01	0.68
BDI-17	1.57 (0.19)	8.18	0.63
F2			
BDI-12	1.00		0.49
BDI-13	1.10 (0.15)	7.15	0.52
BDI-15	1.12 (0.15)	7.42	0.62
BDI-16	1.01 (0.15)	6.92	0.46
BDI-18	0.64 (0.12)	5.41	0.38
BDI-19	1.27 (0.16)	7.87	0.62
BDI-20	1.25 (0.15)	8.29	0.66
F3			
BDI-4	1.00		0.50
BDI-5	0.74 (0.12)	6.23	0.47
BDI-6	1.21 (0.18)	6.65	0.53
BDI-8	1.34 (0.21)	6.41	0.52
BDI-10	1.27 (0.21)	6.15	0.45
BDI-21	0.71 (0.14)	5.05	0.30

Note. *All path coefficients are significant at 5% level of significance.

With respect to the latent factors components of the model (figure 1), the results show that all the three factors have statistically significant associations between them (Negative attitude and Performance difficulty are associated at $r = 0.76$, p-value < 0.001; Negative attitude and Somatic complaints are associated at $r = 0.78$, p-value < 0.001; and Performance difficulty and Somatic complaints are associated at $r = 0.83$, p-value < 0.001). The high correlation between the latent factors is suggestive of the presence of a higher/second-order general factor ("Depression") (Byrne & Baron, 1993; Byrne et al., 1995).

4.2.2.3 *CFA 2*

As indicated in table 3, testing of the hypothesised model for Group 2 (i.e., the validation sample) once again yielded a very well-fitting model (S-B $\chi^2 =$ 546.89, df = 390, $p < 0.0001$, $\chi^2/\mathrm{df} = 1.40$, B-B NNFI = 0.93, CFI = 0.93, SRMR = 0.05, RMSEA = 0.03, with a 90% CI for RMSEA of 0.02- 0.04), and all parameters were statistically significant. From these results, we concluded that the hypothesized model of the BDI–II structure, as shown in figure 1, represented data for nonclinical South African students adequately.

4.2.3 *Stage 2 analyses: MI across race and gender*

Testing for factorial invariance across gender and race was conducted within the framework of MCFA modeling using procedures similar to those outlined in the BDI-II MI literature (cf. Byrne, 2006; Byrne & Stewart, 2006; Byrne et al., 2007; Chen, Sousa, & West, 2005; Whisman, Whiteford, & Gelhorn, 2013; Wu, 2009, 2010). The analyses in the present study can be considered unique in at least two ways. First, unlike most tests for invariance based on the analysis of covariance structures (COVS) only, the present study is based on the analysis of MACS, which allowed us to address the issue of scalar equivalence in testing for differences in the levels or means of the lower-order factors.

Secondly, analyses were based on the EQS 6.1 (Bentler, 2005) program, the only SEM program to date that is capable of yielding corrected goodness-of-fit indices and standard errors in the face of data that are both non-normally distributed and missing completely at random. These distinctive analytic features, although imperative in applications of CFA conducted within the framework of SEM, would appear to be seldom implemented and consequently, infrequently reported in the psychological literature (Byrne et al., 2007).

A series of baseline CFA models of the hypothesized three-factor structure of the BDI-II, established in the previous stages and informed by past research on obtained factor structures in college student samples was first conducted. These results were used to identify a well-fitting model to use in the analyses of factorial invariance. Preliminary models converged on the same lower-order latent factor structure obtained in previous studies (e.g., Byrne & Stewart, 2006; Byrne, Stewart, Kennard, & Lee, 2007). Once this baseline model was shown to be consistent with the data, the analyses then proceeded to test the equivalence of this model across subgroups (i.e., blacks and whites, and males and females), using a series of ordered steps based on integrating approaches outlined for lower-order factor models by Byrne (2006) and by Chen et al. (2005).

The first model specified simply configural invariance, meaning that the same factor structure was estimated simultaneously in both groups but no between-group constraints were placed on the parameter estimates (model 1). Assuming this model is consistent with the data, we advanced the analysis by imposing a series of more stringent between-group constraints to examine factorial invariance. Consistent with both Byrne (2006) and Chen et al. (2005), model 2 was then estimated in which the lower-order loadings were constrained to be equal across groups. This model specifies what is usually meant by MI, allowing differences in factor variances and error variances, but forcing measurement equivalence (equal loadings) across groups. Given equal between-group loadings, if this model remains consistent with the data, then it suggests not only MI but also equivalent between-group variance in the latent factors or traits measured by the items.

Evaluation of model fit was based on multiple criteria that took substantive, statistical and practical fit into account (e.g., Fan & Sivo, 2005; Kim, 2005; Marsh, Hau, & Wen, 2004). First, because the BDI-II data were non-normally distributed, the Y-B χ^2 (Yuan & Bentler, 2000), rather than the uncorrected ML χ^2 statistic is reported. The Y-B χ^2 incorporates a scaling correction for the χ^2 when distributional assumptions are violated. Similar to the χ^2 statistic, use of the Y-B χ^2 is sensitive to sample size. Consequently, other goodness-of-fit statistics, developed and recommended in reporting results for analyses of MACS and MI were also included. These included the CFI, SRMR, and RMSEA and its 90% CI (Hu & Bentler, 1999). CFI values range from 0 to 1.00, with values > 0.95 generally accepted as a good fit. SRMR values range from 0 to 1.00, with values < 0.08 indicating a well-fitting model. The RMSEA is expressed per degree of freedom, which makes it sensitive to model complexity; values < 0.05 indicate acceptable fit. Moreover, probability levels of the equality constraints as determined by the LM test were used for the assessment of race and gender invariance. Equality constraints with $p < 0.05$ were deemed untenable.

Having identified the indices and cut-points used here in the assessment of model fit, it is important to emphasize that these criteria should not be considered a template for the testing of all CFA models as Hu and Bentler (1999) have clearly shown that these fit indices can and do behave differently with diverse sample sizes, estimators, and degrees of model misspecification (see also Marsh et al., 2004). Hence, their recommendations are intended purely as a guide, rather than as black-and-white criteria to use in judging the extent to which hypothesized models fit sample data (Byrne et al., 2007). For both CFI and RMSEA, the robust versions of these measures are reported (*CFI and *RMSEA).

4.2.3.1 *The baseline models*

Tests of the hypothesised BDI-II hierarchical structure (see figure 1) revealed an excellent fit to the data for both blacks and whites (Y-B $\chi^2_{[369]}$ = 448.66; *CFI = 0.96; SRMR = 0.04; *RMSEA = 0.02, with 90% CI = 0.01 to 0.02) and an acceptable fit for males and females (Y-B $\chi^2_{[367]}$ = 526.42; *CFI = 0.93; SRMR = 0.05; *RMSEA = 0.03, with 90% CI = 0.03 to 0.04). All parameter estimates were viable and statistically significant.

4.2.3.2 *Tests for factorial invariance*

Four multigroup models were tested across black and white, and male and female university students, each representing an increasingly more restricted parameterization than its predecessor. As such, these models are said to be hierarchically nested. Results from the related tests for invariance are summarized in tables 5 and 6.

Table 5:

Goodness-of-fit statistics for tests for invariance of BDI–II hierarchical structure for blacks and whites

	Model and constraints	*Y-B χ^2*	*df*	**CFI*	*SRMR*	**RMSEA*	*90% *RMSEA CI*	*Model comparison*	*Δ*CFI*	*Δ*RMSEA*
1.	Configural invariance	448.66	369	0.96	0.04	0.02	0.01, 0.03			
2.	Lower-order factor loadings invariant	469.59	379	0.96	0.05	0.02	0.02, 0.03	2 vs. 1	0.00	0.00
3.	Intercepts invariant	633.77	392	0.95	0.05	0.04	0.03, 0.04	3 vs. 1	-0.01	0.02
4.	Latent factor means invariant	597.98	389	0.95	0.06	0.04	0.03, 0.04	4 vs. 1	-0.01	0.02

Note. *p* < .001; df = degrees of freedom; Y-B χ^2 = Yuan-Bentler scaled Chi-square test; *CFI = robust Comparative Fit Index; *RMSEA = robust root mean square error of approximation and its 90% confidence interval; SRMR = Standardized Root Mean-Square Residual; Δ*CFI = Comparative Fit Index difference value; Δ*RMSEA = robust root mean square error of approximation difference value.

Table 6:

Goodness-of-Fit Statistics for Tests of Invariance of BDI–II Hierarchical Structure for Males and Females

	Model and constraints	*Y-B* χ^2	*df*	**CFI*	*SRMR*	**RMSEA*	*90% *RMSEA CI*	*Model comparison*	*Δ*CFI*	*Δ*RMSEA*
1.	Configural invariance	526.42	367	0.93	0.05	0.03	0.02, 0.04			
2.	Lower-order factor loadings invariant	555.18	385	0.93	0.06	0.03	0.03, 0.04	2 vs. 1	0.00	0.00
3.	Intercepts invariant	617.97	406	0.93	0.06	0.03	0.03, 0.04	3 vs. 1	0.00	0.00
4.	Latent factor means invariant	593.07	403	0.93	0.06	0.03	0.03, 0.04	4 vs. 1	0.00	0.00

Note. $p < .001$; df = degrees of freedom; Y-B χ^2 = Yuan-Bentler scaled Chi-square test; *CFI = robust Comparative Fit Index; *RMSEA = robust root mean square error of approximation and its 90% confidence interval; SRMR = Standardized Root Mean-Square Residual; Δ*CFI = Comparative Fit Index difference value; Δ*RMSEA = robust root mean square error of approximation difference value.

Nested models can be compared in pairs by calculating the differences in their overall χ^2 values and the related degrees of freedom; the χ^2 difference value ($\Delta\chi^2$) is distributed as χ^2, with the degrees of freedom equal to the difference in degrees of freedom (Δdf). Analogously, the same comparisons can be based on the Y-B χ^2, except that a correction to this difference value is needed because it is not distributed as χ^2 (Bentler, 2005). Historically, evidence in support of invariance has been based on the $\Delta\chi^2$ test. If this value is statistically significant, it suggests that the constraints specified in the more restrictive model do not hold (i.e., the two models are not equivalent across groups).

Recently, however, researchers (e.g., Cheung & Rensvold, 2002; Little, 1997; Marsh, Hey, & Roche, 1997) have argued that this $\Delta\chi^2$ value is as sensitive to sample size and non-normality as the χ^2 statistic itself, thereby rendering it an impractical and unrealistic criterion on which to base evidence of invariance. As a consequence, there has been an increasing tendency to argue for evidence of invariance based on two alternative criteria: (a) the multigroup model exhibits an adequate fit to the data, and (b) the CFI (or Δ*CFI) values between models is negligible. Until the recent simulation research of Cheung and Rensvold (2002), use of the Δ*CFI difference value has been of a purely exploratory nature.

In contrast, based on the examination of properties related to 20 goodness-of-fit indices within the context of invariance testing, Cheung and Rensvold (2002) recommended that the Δ*CFI provides the best information in determining evidence of MI and suggested that its difference value should not exceed 0.01. However, more recently, Chen (2007) conducted simulation studies to examine the performance of various relative fit measures in examining MI in groups with large sample sizes. As a result of these studies, she made recommendations for particular measures of relative fit, and appropriate cut-off values, which have proven to be informative for examining

MI in large samples. In particular, she recommended that MI in larger samples should be rejected when ΔCFI ≥ 0.01 and when ΔRMSEA ≥ 0.015. In lieu of this statistically based (as opposed to heuristically based) research, the current work builds conclusions regarding invariance on these two measures of relative fit and adopts Cheung and Rensvold (2002) and Chen's (2007) cut-off values for rejecting MI based on their practical approach, although we base our analyses on the robust versions of these measures (Δ*CFI and Δ*RMSEA).

In reviewing the results in table 5, we see that model 1, the configural model in which no equality constraints were imposed, represented an excellent fit to the data, indicating that blacks and whites have the same basic conceptualization of depression. This model serves as the baseline against which all remaining models are compared in the process of determining evidence of invariance. Model 2 (metric invariance), in which all lower-order factor loadings were equally constrained, also represented an excellent fit to the model with a resulting Δ*CFI and Δ*RMSEA values of 0.00. Nevertheless, 3 constraints associated with lower-order factor loadings were found to be untenable across race. These constrained parameters, and their associated univariate Mχ^2 and probability values are summarized in table 7.

Each Mχ^2 value represents the expected reduction in the overall model fit χ^2 value if the related equality constraints were released. Of the three non-invariant items shown in table 7 for model 2, those measuring concentration, guilt and pessimism appear to be discrepant across black and white university students. In an effort to uncover a reason as to why the first three items should be differentially valid across race, we examined their factor loadings, distributional statistics and response frequencies. Several interesting comparisons can be made from the information presented in table 8. For instance, we can observe a difference between blacks and whites with respect to the size of the factor loading, the degree of skewness and kurtosis and distributional characteristics for item 19 (concentration). Because factor loadings reflect the extent to which a respondent endorses an item, it is apparent that concentration is of more concern to white university students than it is so for blacks, although loadings in both groups are significant.

Responses to Items 5 (guilt) and 2 (pessimism) suggest that, in addition to differential endorsement, distributional characteristics related to these items possibly accounted for the noninvariance finding. For Item 5, we note that in addition to guilt being more salient for blacks (0.50 vs 0.44) more whites than blacks (20.7%) reported no guilt feelings (the ‘0’ category). In contrast, more blacks than whites (19.5%) indicated feelings of guilt (the ‘1’ category). Likewise, for Item 2, endorsement was stronger for blacks than for whites (0.42 vs 0.39). Furthermore, while more blacks (18.2%) reported not feeling pessimistic about their future (the ‘0’ category), more whites than blacks (18.5%) responded that they felt more discouraged about their future (the ‘1’ category). The satisfaction of metric invariance for the remaining eighteen items implies that these items of BDI-II have equal salience for blacks and whites.

Model 3 (scalar invariance) additionally constrains the intercepts to be equal between groups, thus forcing equality of the variances/covariance matrices between the races. Once again, results yielded an excellent fit and a Δ*CFI value of 0.01. Likewise, for the final remaining model tested, the multigroup model retains the same excellent fit and the Δ*CFI value never exceeds 0.01. However, scalar invariance was achieved by removing two constraints of intercepts (item 5 and 14) (see table 11). The findings of two noninvariant intercepts implied that there is differential additive response style (ARS) bias (Cheung & Rensvold, 2000) for the BDI-II across race groups. That is, whites systematically endorse higher item responses in item 5 and 14. Consistent with Cheung and Rensvold’s (2002) and Chen’s (2007) recommendations, we consider this model to exhibit evidence of invariance across the two race groups.

Similarly, for the models on gender tested in table 6, the multigroup models retain the same acceptable fit with the Δ*CFI and Δ*RMSEA values never exceeding 0.01. However, 3 constraints associated with lower-order factor loadings were also found to be untenable across gender. These constrained parameters, and their associated univariate Mχ^2 and probability values are summarized in table 9. Of the three non-invariant items shown in table 9 for model 2, those measuring loss of energy, agitation and worthlessness appear to be discrepant across gender.

In an effort to uncover a reason as to why the first three items should be differentially valid across race, we examined their factor loadings, distributional statistics and response frequencies. This information is summarized in table 10 and several interesting comparisons can be made from the information presented. For instance, we can observe a difference between males and females with respect to the distributional characteristics for item 15 (loss of energy). Given the relatively similar pattern of the endorsement of item content (0.56 vs. 0.57), it seems likely that findings of gender non-invariance are a result of differential item responses. For example, more males than females (19.2%) reported no loss of energy (the ‘0’ category). In contrast, more females than males (13%) reported having less energy (the ‘1’ category). Likewise, for item 14, item content endorsement was equivalent for both males and females (0.57 vs. 0.60). However, more males (9.2%) than females reported no feelings of worthlessness (the ‘0’ category), whereas more females (4.1%) than males reported some feelings of worthlessness. Finally, given the relatively similar pattern of responses by males and females to Item 11 (agitation), it seems likely that findings of gender non-invariance derived from the differential endorsement of item content, with the item being more salient for females than males (0.51 vs. 0.40).

Additionally, scalar invariance was achieved for gender by removing three constraints of intercepts (items 11, 14 and 18). Of the three non-invariant intercepts identified in model 3 presented in table 12, females expressed higher response scores on item 11 (agitation), 14 (worthlessness) while males endorsed higher item responses on item 18 (appetite). The findings of three noninvariant intercepts implied that there is ARS bias (Cheung & Rensvold, 2000) for the BDI-II across gender groups. That is, females systematically endorse higher item responses in items 11 and 14.From the findings based on these criteria, then, we concluded that the lower-order BDI-II structure, as portrayed in Figure 1, was operating equivalently across race and gender. Furthermore, the fit of the models never deteriorated when we assumed equal variance in the latent factors between the two groups.

Table 7:

Summary statistics for noninvariant parameters across race

Constrained parameter	*LM*χ^2	*Probability*
Lower-order factor loadings		
Item 19 (concentration) on Factor 2	21.28	0.000
Item 5 (guilt) on Factor 3	5.17	0.023
Item 2 (pessimism) on Factor 1	4.99	0.026

Table 8:

Item statistics for noninvariant items across race

		Sample distribution[b]		Frequency distribution[c]			
Item	*Factor loading*[a]	*SK*	*KU*	*0*	*1*	*2*	*3*
Blacks							
Item 19	0.41	0.31	-1.19	31.6	31.3	27.6	1.6
Item 5	0.50	0.32	-0.38	40.8	53.9	5.2	0.2
Item 2	0.42	3.20	11.71	86.2	11.4	1.7	0.7
Whites							
Item 19	0.46	1.06	0.66	55.0	36.4	7.6	1.0
Item 5	0.44	1.34	2.01	61.5	34.4	3.1	1.0
Item 2	0.39	1.53	2.94	68.0	29.9	1.4	0.7

Note. a = standardized solution; b = SK: skewness, KU: kurtosis; c = in percentages.

Table 9:

Summary statistics for noninvariant parameters across gender

Constrained parameter	*LM*χ^2	*Probability*
Lower-order factor loadings		
Item 15 (loss of energy) on Factor 3	8.61	0.003
Item 11 (agitation) on Factor 1	6.74	0.009
Item 14 (worthlessness) on Factor 1	7.15	0.007

Table 10:

Item statistics for noninvariant items across gender

Item	*Factor loading*[a]	*Sample distribution*[b]		*Feguency distribution*[c]			
		SK	*KU*	*0*	*1*	*2*	*3*
Males							
Item 15	0.56	0.85	-0.28	57.9	36.8	5.4	-
Item 11	0.40	1.85	2.75	66.3	24.5	3.1	6.1
Item 14	0.57	2.90	7.30	87.7	7.3	5.0	-
Females							
Item 15	0.57	0.95	0.28	38.7	49.8	9.9	1.5
Item 11	0.51	1.81	3.56	56.5	34.8	3.7	5.0
Item 14	0.60	2.73	6.53	78.5	11.4	9.1	0.9

Note. a = standardized solution; b = SK: skewness, KU: kurtosis; c = in percentages.

Table 11:

The intercepts of noninvariant items in Model 3 across race

	Factor	*Estimated values*	
		Blacks	*Whites*
Item 5 (guilt)	Somatic complaints	0.51	0.47
Item 14 (worthlessness)	Negative attitude	0.57	0.67

Table 12:

The intercepts of noninvariant items in Model 3 across gender

		Estimated values	
	Factor	*Males*	*Females*
Item 11 (agitation)	Negative attitude	0.40	0.51
Item 14 (worthlessness)	Negative attitude	0.58	0.60
Item 18 (appetite)	Somatic complaints	0.32	0.32

4.2.3.3 Latent mean differences test

Previous studies on group differences in depression using the BDI-II reflect several primary limitations. Most of this past studies had conducted mean differences without testing for MI or providing established evidences of MI, thus risking comparisons being meaningless because of measurement bias. Establishing MI of the BDI-II is critical to compare groups in terms of their trait scores. This means that when MI is established, measurement bias associated with groups is absent; as such, latent variables or factors are scaled equally (Meredith, 1993; Meredith & Horn, 2001). Based on empirical findings (e.g., Byrne et al., 1993; Osman et al., 2004; Santor et al., 1994; Wu, 2009, 2010), it could be asserted that, if there is no evidence of MI of the BDI-II across groups, the basis for making an inference is lacking—the findings of differences between groups would be ambiguous and might be interpreted misleadingly.

Furthermore, studies have mostly focused on group differences at the level of the overall depression rather than on specific factors/dimensions. Only Osman et al. (1997) conducted analyses of gender differences on total scores in addition to the factor scales. In their study, a three-factor model (i.e., Negative attitude, Performance difficulty, and Somatic elements) was best fitted to the data of the BDI-II. Their subsequent analyses showed that females reported higher scores in terms of their overall depression (i.e., total BDI-II scores) as well as on Negative attitude and Performance difficulty factors. Yet no significant gender difference was observed for the Somatic element factor. Based on the study by Osman et al (1997), it could be posited that an examination of gender differences only at the level of overall depression may leave open questions concerning whether differences on particular factors contributed to differences in overall depression scores; or whether gender differences, if they were not revealed in total scores could, however, still be identified on particular factors (Wu, 2010).

Lastly, research has also over-emphasized observed mean differences instead of latent mean differences, thereby running the risk of ignoring the impact of measurement error (Wu, 2010). The previous studies referred to mostly used *t* test analysis or similar statistical methods to compare groups in terms of observed mean scores. When a significant difference in observed mean scores is found, this is taken to mean evidence for group differences. However, observed scores on a factor confound factor and indicator (item) variance, which in turn may lead to conclusions of an erroneous nature. That is, the observed mean differences could be masked depending on the degree of confounding. In contrast to observed mean scores, latent mean scores are estimated by partialing out variance attributable to measurement error (Cheung & Rensvold, 2000). As such, comparing latent mean differences allows for measurement errors to be avoided. When comparing latent mean differences, scalar invariance (strong MI) is required (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000). Modeling latent means in SEM could address the limitations mentioned above (e.g., Cheung & Rensvold, 2000; Ployhart & Oswald, 2004; Wu, 2010).

Given the support for configural, metric and scalar invariance in the present study, a comparison of latent factor mean differences across race and gender groups was possible. MCFA may also be used to test whether the latent factor means differ across the groups. In a usual covariance structure model (Hoyle, 1991), the covariance matrix is computed from deviation scores so that the means of all measured variables will be zero. As a result, the means of all latent constructs are assumed to be zero. To test the latent construct mean differences, a combined mean and covariance structure model must be used (Bentler, 1989; Bollen, 1989; Sörbom, 1978). To estimate the difference between the factor means, one group is usually chosen as a reference or baseline group and its latent means are set to zero. The latent means of the other group, which actually represent the difference between the factor means in the two groups, are estimated. The significance test (Wald or *z* test) for the latent means of the second group provides a test for significance of the difference between the means of the two groups on the latent construct (Aiken, Stein, & Bentler, 1994).

Accordingly, latent mean value was set to zero in the black and male group and freely estimated for whites and females. As seen in table 13, there were no significant latent mean and observed mean differences on the Negative attitude factor (z = 0.23, $p > .05$) between the race groups. However, latent mean and observed mean differences on the Performance difficulty factor (z = 4.16, $p < .05$) and the Somatic complaints factor (z = 3.74, $p < .05$) were significant, with blacks endorsing higher scores.

Table 14 shows that, there were significant latent mean and observed mean differences on the Negative attitude factor (z = 3.95, $p < .05$) and the Somatic complaints factor (z = 4.63, $p < .05$), with females endorsing higher scores. However, latent mean differences on the Performance difficulty factor (z = 1.27, $p > .05$) were not significant, although observed mean differences were significant. Even so, results also showed that females have significantly higher observed scores on the Performance difficulty factor than males even when conducting one-tailed hypothesis testing. Thus, one may conclude that there were gender differences on this factor based on observed mean differences.

Table 13:

Differences between blacks and whites on latent constructs

	Latent mean analyses	*Observed mean analyses*					
		Blacks		*Whites*			
Dimension	*Model 3 z*	*M*	*SD*	*M*	*SD*	*t*	*d*
Negative attitude	0.23	2.76	2.95	2.95	3.00	-0.85	0.06
Performance difficulty	4.16*	5.57	3.26	4.31	2.94	5.53*	0.41
Somatic complaints	3.74*	4.13	3.22	2.30	2.31	8.62*	0.66

Note. z = Wald significance test; *p <.05; d = effect size (Cohen's d).

Table 14:

Differences between males and females on latent constructs

	Latent mean analyses	*Observed mean analyses*					
		Males		*Females*			
Dimension	*Model 3 z*	*M*	*SD*	*M*	*SD*	*t*	*d*
Negative attitude	3.95*	2.35	2.63	3.04	3.09	-3.16*	0.24
Performance difficulty	1.27	4.40	3.03	5.46	3.23	-4.57*	0.34
Somatic complaints	4.63*	3.16	3.03	3.61	3.06	-1.99*	0.15

Note. z = Wald significance test; *p <.05; d = effect size (Cohen's d).

4.2.4 *Stage 3 analyses: Longitudinal MI of the hypothesized three-factor model*

In addition to estimating the internal consistency of the factors themselves, it was considered important also to assess the stability of hierarchically structured factor loadings across time. To this end, the invariance of all factor loadings estimated at time 1 across the time lag of 2 weeks was tested. As such, the overall fit of the 3-factor model at time 1 and again at time 2, with no equality constraints imposed, was firstly tested. Although the fit of this model was less than the cut-off point suggested by Hu and Bentler (1999) (*CFI = 0.94, table 15), it nevertheless was indicative of an adequately fitting model that included two time points of data. Following this, a test for the invariance of the lower-order factor loadings was conducted. As such, the three-factor model (time 1 and time 2) was again estimated, but this time with equality constraints placed on all lower-order factor loadings across time 1 and time 2.

Determining evidence of invariance involves testing and comparing the difference in fit for a series of nested models. For example, comparison of a model in which no constraints are imposed (model 1: configural invariance) with one in which equality constraints are specified for all lower-order factor loadings (model 2: metric invariance) would constitute a nested model comparison. Based on Cheung and Rensvold (2002) and Chen's (2007) proposal that the difference in CFI and RMSEA values (Δ*CFI and Δ*RMSEA) equal to or less than 0.01 can rightfully serve as viable evidence of invariance and judgement of the overall models exhibiting adequate fit, the hierarchical structure *CFIs and *RMSEAs were compared.

Table 15:

Goodness-of-Fit and Comparative Statistics for Tests for Invariance of BDI–II Hierarchical Structure across Time

	Model and constraints	*Y-B χ^2*	*df*	**CFI*	*SRMR*	**RMSEA*	*90% *RMSEA CI*	*Model comparison*	*Δ*CFI*	*Δ*RMSEA*
1.	Configural invariance	550.25	365	0.94	0.04	0.03	0.02, 0.03			
2.	Lower-order factor loadings invariant	567.23	382	0.94	0.05	0.03	0.02, 0.03	2 vs. 1	0.00	0.00
3.	Intercepts invariant	625.11	399	0.94	0.06	0.03	0.03, 0.04	3 vs. 1	0.00	0.00
4.	Latent factor means invariant	587.28	396	0.94	0.05	0.03	0.02, 0.03	4 vs. 1	0.00	0.00

Note. p < .001; df = degrees of freedom; Y-B χ^2 = Yuan-Bentler scaled Chi-square test; *CFI = robust Comparative Fit Index; *RMSEA = robust root mean square error of approximation and its 90% confidence interval; SRMR = Standardized Root Mean-Square Residual; Δ*CFI = Comparative Fit Index difference value; Δ*RMSEA = robust root mean square error of approximation difference value.

Accordingly, given an overall adequate fit of 0.94 and no deterioration in the overall fit (i.e., when we assume equal variance in the latent factors and intercepts) between models 1, 2, 3 and 4 (Δ*CFI = 0.00 and Δ*RMSEA = 0.00), it can be concluded that the lower-order factor loadings were invariant across time. In Model 3, extremely stringent assessment of invariance, we tested for the equality of intercepts across time; this test measures scalar invariance of the BDI-II. Once again, results yielded a Δ*CFI of 0.00 and Δ*RMSEA of 0.00, thereby providing credible evidence of invariance across time.

4.2.5 Stage 4 analyses: Internal consistency reliability for the total scale and the three lower order factors

Internal consistency reliability coefficients, as computed for Cronbach's coefficient alpha, are reported in table 16 for both Time 1 and Time 2. Internal consistency of the total scale score for overall depression was high at Time 1 (α = 0.84), and even slightly higher at Time 2 (α = 0.90). Although internal consistency for the Somatic complaints subscale was somewhat weaker than for the Negative attitude and Performance difficulty subscales, it nonetheless exhibited adequate reliability.

Byrne et al. (2004) explain the occurrence of this relatively lower alpha coefficient for the Somatic complaints factor as not necessarily an anomaly, but due to the general relative rarity of somatic symptoms in the normal population.

Table 16:

Internal consistency of the BDI-II across time

	Internal consistency coefficients	
BDI-II Subscales	*Time 1*	*Time 2*
Negative attitude	0.73	0.82
Performance difficulty	0.70	0.78
Somatic complaints	0.62	0.75
Depression (total score)	0.84	0.90

Note. Time 1 N = 919 and Time 2 N = 304.

4.2.6 *Stage 5 analyses: Correlations between depression and theoretically linked external criteria*

4.2.6.1 **Convergent and discriminant validity**

In table 17 correlations between the BDI-II and the HSCL-10 (anxiety), and HSCL-15 (depression) are listed. The correlation between the BDI-II and HSCL-15 (depression) was expected to yield convergent validity, and discriminant validity with HSCL-10 (anxiety). Correlation analysis shows that the BDI-II is statistically significantly and positively associated with both the HSCL-15 (depression) ($r = 0.67$, $p < .01$) and the HSCL-10 (anxiety) ($r = 0.50$, $p < .01$).

Table 17:

Correlations between the BDI-II and HSCL-25 subscales, full sample (N = 919)

	BDI-II	*HSCL-10 (anxiety)*
BDI-II	-	
HSCL-10 (anxiety)	0.50**	-
HSCL-15 (depression)	0.67**	0.64**

Note. ** $p < .01$.

Likewise, a between (or across) methods convergent validation or cross validation was explored. The BDI-II scale scores were compared with the results of SCID-I/NP of a random subsample (n = 20) of the student participants. The two distinct methods (i.e., self-report scale and clinical interview) of measuring depression were found to be congruent and yielded comparable data. All the 20 participants were found to be asymptomatic on the BDI-II and SCID-I/NP (depression subscale), and this provided evidence of convergent validity.

The interrater reliability assessed by Cohen's kappa to measure agreements between the BDI-II scores and SCID-I diagnosis were also planned. However, due to the small sample size (i.e., 20 participants) and that all classifications of participants on the instruments were constant (i.e., all were not depressed), it was not theoretically justifiable or necessary to calculate the kappa.

4.2.6.2 **Construct validity**

As indicated earlier, in the literature review, scores on depression should correlate with risk factors and environmental concomitants and precipitants of depression (e.g., stressful life events). For this reason, it was anticipated that hopelessness and stressful life events would display strong correlations with depression scores (Byrne et al., 2007). Results presented in table 18 seem to confirm the expectations quite adequately. Correlation analysis shows that depression correlates statistically significantly and positively with hopelessness ($r = 0.44$, $p < .01$), perceived stress ($r = 0.28$, $p < .01$), and self-efficacy ($r = 0.39$, $p < .01$), and statistically significantly but negatively with self-esteem ($r = -0.51$, $p < .01$).

Table 18:

Correlation between the BDI-II and external criteria, full sample (N = 919)

External Criteria	*BDI-II*
BHS	0.44**
PSS-10	0.28**
SGSES	0.39**
RSES	-0.51**

Note. ** $p < .01$.

4.2.7 *Stage 6 analyses: Examination of Byrne et al.'s (2004) three-factor lower-order model*

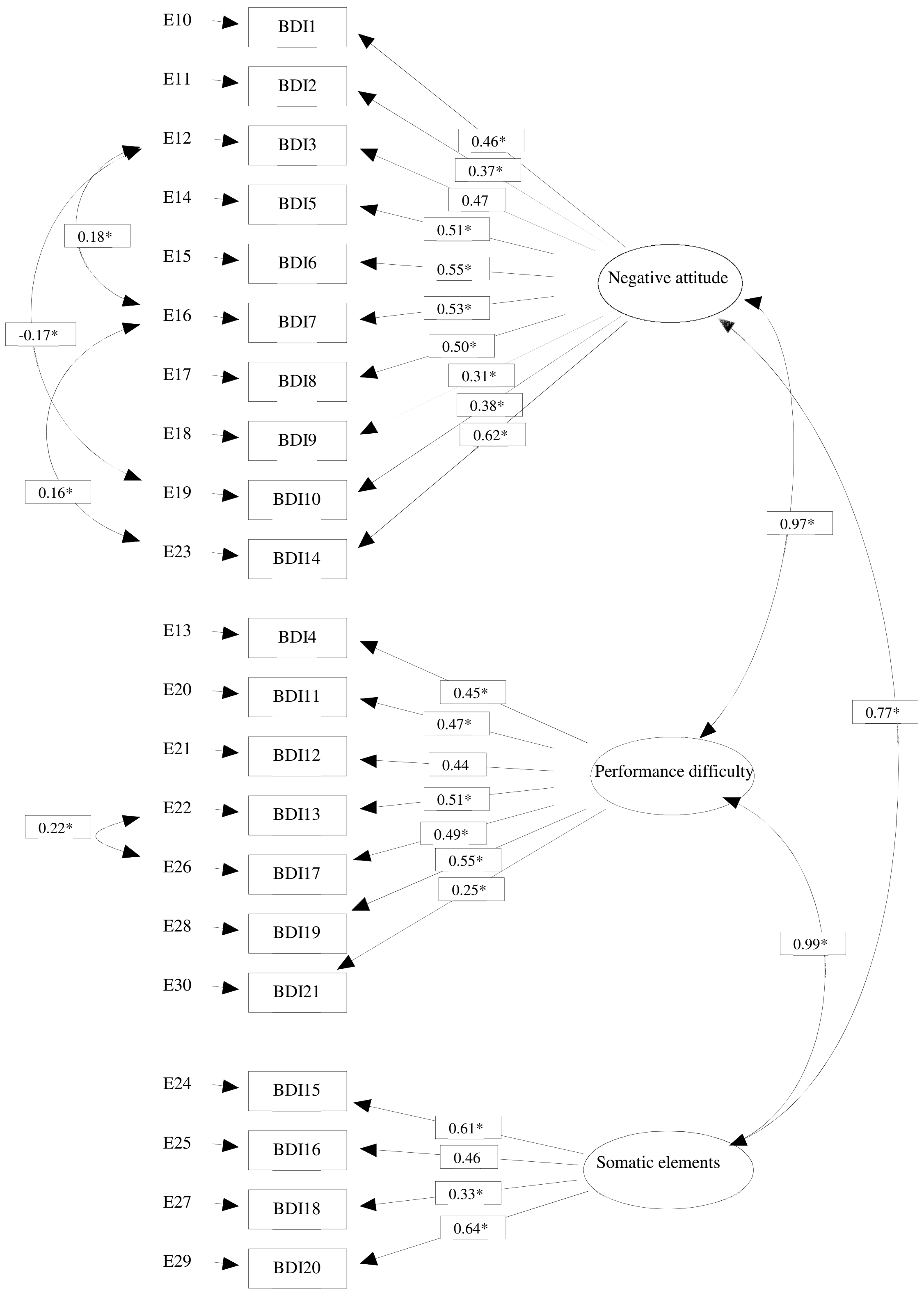
Byrne et al.'s (2004) three-factor lower-order model (see figure 2) was also tested and cross-validated for two groups formed by randomly splitting the data into two nearly equal groups using CFA. Moreover, MI for the model was tested for race, gender and across time.

4.2.7.1 *CFA 1*

Based again on data from Group 1 (n = 460), the calibration sample, the validity of the BDI–II three-factor structure reported by Byrne and colleagues (Byrne, Baron & Campbell, 1993, 1994; Byrne, Baron, Larsson & Melin, 1995, 1996, 1998; Byrne & Campbell, 1999, Byrne et al., 2004) was tested. Results revealed a well-fitting model to the data. Goodness-of-fit statistics related to the test of this hypothesized model are shown in table 19.

Figure 2:

Three-factor model (Byrne et al., 2004)



Note. The values on the path diagram are standardized regression coefficients, with arrows pointing from latent variables to the observed variables.

Table 19:

Hypothesized model of BDI-II structure: Goodness-of-fit statistics

Group	*df*	*B-B NNFI*	*S-B χ^2*	**CFI*	**RMSEA*	*90%*RMSEA CI*	*SRMR*
Calibration[a]	182	0.91	272.22	0.93	0.03	0.02, 0.04	0.05
Validation[b]	182	0.92	264.12	0.93	0.03	0.02, 0.04	0.05

Note. [a]n = 460, [b]n = 459; df = degrees of freedom; B-B NNFI = Bentler-Bonett Non-Normed Fit Index; S-B χ^2 = Satorra-Bentler Scaled Chi-square test; *CFI = robust Comparative Fit Index; *RMSEA = robust root mean square error of approximation and its 90% confidence interval; SRMR = Standardized Root Mean-Square Residual.

Byrne et al.'s (2004) hypothesized model (i.e. three factors model) was tested within the SEM framework. For the initial structural model, the significant χ^2 value and its ratio to df is greater than 1.5 (S-B $\chi^2 = 311.42$, df = 186, $p < 0.0001$, $\chi^2/\mathrm{df} = 1.67$), along with poor fit indices, namely, B-B NNFI = 0.88 and CFI = 0.90, but acceptable SRMR = 0.05 and RMSEA = 0.04 with a 90% CI for RMSEA of 0.03 to 0.05, implied that the initial model did not fit the data well. Therefore, we modified the model by adding paths based on the LM tests. The LM tests suggested the inclusion of 4 correlations between measurement errors out of 210 possible correlations. The final model for the study is presented in figure 2. The fit statistics for the final model in figure 1 (the ratio of S-B χ^2 and its df [272.207/182 = 1.49] is less than 1.5, CFI = 0.93 < 0.95, B-B NNFI = 0.91 > 0.90, SRMS = 0.05 and RMSEA 0.03 with 90% CI for RMSEA of 0.02 to 0.04), are acceptable as they agree with the cut-off values suggested by the literature.

The parameter estimates of the final SEM (figure 2) are shown in table 20. All path coefficient estimates have the expected signs. The magnitudes of the standardized path coefficient estimates, for the measurement components of the model, suggest that the items BDI-5, BDI-6, BDI-7, BDI-8 and BDI-14 have a stronger effect on Factor 1: "Negative attitude" than BDI-1, BDI-2, BDI-3, BDI-9 and BDI-4; BDI-11, BDI-13, BDI-17 and BDI-19 have a stronger effect on Factor 2: "Performance difficulty" than BDI-12 and BDI-21; and BDI-15, BDI-16, and BDI-20 have a stronger effect on Factor 3: "Somatic elements" than BDI-18.

Table 20:

Maximum likelihood estimates of measurement and structural path coefficients for the calibration data (n = 460)

Items	*Coefficient (Robust SE)*	*Robust t – ratio**	*Standardized coefficient*
<u>F1</u>			
BDI-1	1.00(0.20)	5.12	0.46
BDI-2	0.66(0.13)	5.15	0.37
BDI-3	0.96(0.15)		0.47
BDI-5	1.00		0.51
BDI-6	1.84(0.29)	6.30	0.55
BDI-7	1.34(0.18)	7.66	0.53
BDI-8	1.79(0.33)	5.48	0.50
BDI-9	1.43(0.10)	4.22	0.31
BDI-10	1.51(0.33)	4.58	0.62
BDI-14			
<u>F2</u>			
BDI-4	0.97(0.14)	6.77	0.45
BDI-11	1.14(0.17)	6.63	0.48
BDI-12	1.00		0.44
BDI-13	1.21(0.17)	7.06	0.51
BDI-17	0.98(0.15)	6.31	0.49
BDI-19	1.34(0.18)	7.39	0.55
BDI-21	0.62(0.15)	4.15	0.25
<u>F3</u>			
BDI-15	1.11(0.14)	7.86	0.61
BDI-16	1.00		0.46
BDI-18	0.60(0.11)	5.64	0.33
BDI-20	1.19(0.14)	8.31	0.64

Note. All path coefficients are significant at 5% level of significance.

With respect to the latent factors components of the model (figure 2), the results show that all the three factors have statistically significant associations between them (Negative attitude and Performance difficulty have $r = 0.97$ with p-value < 0.001; Negative attitude and Somatic elements have $r = 0.77$ with p-value < 0.001; and Performance difficulty and Somatic elements, $r = 0.99$ with p-value < 0.001).

4.2.7.2 *CFA 2*

As indicated in table 19, the testing of Byrne et al.'s (2004) model for Group 2 (i.e., the validation sample) once again yielded a well-fitting model, and all parameters were statistically significant. From these results, we concluded that the hypothesized model of the BDI–II structure, as shown in figure 2, represented data for nonclinical South African students adequately.

4.2.7.3 *Factorial invariance across gender and race*

A series of baseline CFA models of our hypothesized three-factor structure of the BDI-II, established in the previous stages and informed by past research on obtained factor structures in college student samples was first conducted. These results were used to identify a well-fitting model to use in the analysis of factorial invariance. Preliminary models converged on the same lower-order latent factor structure obtained by Byrne and colleagues (Byrne et al., 2004; Byrne et al., 2007; Byrne & Stewart, 2006). Once this baseline model was shown to be consistent with the data, the analyses then proceeded to test the equivalence of this model across subgroups (i.e., blacks and whites, and males and females), using a series of ordered steps based on integrating approaches outlined for lower-order factor models by Byrne (2006) and by Chen et al. (2005). The first model specified simply configural invariance, meaning that the same factor structure was estimated simultaneously in both groups but no between-group constraints were placed on the parameter estimates (model 1). Assuming this model is consistent with the data, we proceeded by imposing a series of more stringent between-group constraints to examine factorial invariance.

Consistent with both Byrne (2006) and Chen et al. (2005), model 2 is then estimated in which the lower-order loadings are constrained to be equal across groups. This model specifies what is usually meant by MI, allowing differences in factor variances and error variances, but forcing measurement equivalence (equal loadings) across groups. Given equal between-group loadings, if this model remains consistent with the data, then it suggests not only MI but also equivalent between-group variance in the latent factors or traits measured by the items.

Several indices for evaluating the model fit were used. First, because the BDI-II data were non-normally distributed, the S-B χ^2 (Satorra & Bentler, 1988) was used instead of the uncorrected maximum likelihood chi-square (χ^2). The S-B χ^2 incorporates a scaling correction for the χ^2 when distributional assumptions are violated. Similar to the χ^2 statistic, use of the S-B χ^2 is sensitive to sample size. Consequently, other goodness-of-fit statistics, developed and recommended in reporting results for analyses of MI, were also included. These included the CFI, SRMR, and RMSEA and its 90% CI (Hu & Bentler, 1999). CFI values range from 0 to 1.00, with values > 0.95 generally accepted as a good fit. SRMR values range from 0 to 1.00, with values < 0.08 indicating a well-fitting model. The RMSEA is expressed per degree of freedom, which makes it sensitive to model complexity; values < 0.05 indicate acceptable fit. For both CFI and RMSEA, the robust versions of these measures are reported (*CFI and *RMSEA).

The baseline models

Tests of the hypothesized BDI-II hierarchical structure (see figure 2) revealed an acceptable fit to the data for both blacks and whites (S-B $\chi^2_{[372]} = 493.67$; *CFI = 0.93; SRMR = 0.05; *RMSEA = 0.03, with 90% CI = 0.02 to 0.04) and, males and females (S-B $\chi^2_{[372]}$ = 588.57; *CFI = 0.91; SRMR = 0.05; *RMSEA = 0.04, with 90% CI = 0.03 to 0.04). All parameter estimates were viable and statistically significant. The asterisks associated with the CFI and RMSEA values indicate robust (i.e., corrected) versions of these indices.

Tests for Factorial Invariance

Two multigroup models were tested across black and white, and male and female university students, each representing an increasingly more restricted parameterization than its predecessor. As such, these models are said to be hierarchically nested. Results from the related tests for invariance are summarized in tables 21 and 22.

Table 21:

Goodness-of-fit statistics for tests for invariance of BDI–II hierarchical structure for blacks and whites

	Model and constraints	*S-B χ^2*	*df*	**CFI*	*SRMR*	**RMSEA*	*90% *RMSEA CI*	*Model comparison*	*Δ*CFI*	*Δ*RMSEA*
1.	Configural invariance	493.67	372	0.93	0.05	0.03	0.02, 0.04			
2.	Lower-order factor loadings invariant	543.91	393	0.93	0.06	0.03	0.02, 0.04	2 vs. 1	0.00	0.00

Note. *p < .001; df = degrees of freedom; S-B χ^2= Satorra-Bentler scaled Chi-square test; *CFI = robust Comparative Fit Index; *RMSEA = robust root mean square error of approximation and its 90% confidence interval; SRMR = standardized root mean-square residual; Δ*CFI = Comparative Fit Index difference value; Δ*RMSEA = robust root mean square error of approximation difference value.

Table 22:

Goodness-of-fit statistics for tests for invariance of BDI–II hierarchical structure for males and females

	Model and constraints	*S-B χ^2*	*df*	**CFI*	*SRMR*	**RMSEA*	*90% *RMSEA CI*	*Model comparison*	*Δ*CFI*	*Δ*RMSEA*
1.	Configural invariance	588.57	372	0.91	0.05	0.04	0.03, 0.04			
2.	Lower-order factor loadings invariant	585.36	389	0.92	0.07	0.03	0.03, 0.04	2 vs. 1	0.01	-0.01

Note. **p* < .001; df = degrees of freedom; S-B χ^2 = Satorra-Bentler scaled Chi-square test; *CFI = robust Comparative Fit Index; *RMSEA = robust root mean square error of approximation and its 90% confidence interval; SRMR = standardized root mean-square residual; Δ*CFI = Comparative Fit Index difference value; Δ*RMSEA = robust root mean square error of approximation difference value.

In reviewing the results in table 21, it can be seen that model 1, the configural model in which no equality constraints were imposed, represented an acceptable fit to the data. This model serves as the baseline against which all remaining models are compared in the process of determining evidence of invariance. Model 2, in which all first-order factor loadings were constrained equal, also represented an excellent fit to the model with a resulting Δ*CFI value of 0.00 and Δ*RMSEA value of 0.00. Consistent with Cheung and Rensvold's (2002) and Chen (2007) recommendations, we consider this model to exhibit evidence of invariance across the two race groups. Likewise, for the models on gender tested in table 22, the multigroup models retain the same excellent fit and the Δ*CFI and Δ*RMSEA values never exceeds 0.01. From the findings based on these criteria, then, we concluded that the lower-order BDI-II structure, as portrayed in figure 2, was operating equivalently across black and white, and male and female South African university students.

4.2.7.4 *Longitudinal MI*

Additionally, it was considered significant also to assess the stability of hierarchically structured factor loadings across time. To this end, the invariance of all factor loadings estimated at time 1 across the time lag of 2 weeks was tested. As such, the overall fit of the three-factor model at time 1 and again at time 2; with no equality constraints were imposed, was firstly tested. The fit of this model (*CFI = 0.89), as shown in table 23, was way less than the cut-off point suggested by Hu and Bentler (1999), and indicative of a poor fitting model that included two time points of data. Following this, a test for the invariance of the lower-order factor loadings was conducted. As such, the three-factor model (time 1 and time 2) was again estimated, but this time with equality constraints placed on all lower order factor loadings across time 1 and time 2.

Determining evidence of invariance involves testing and comparing the difference in fit for a series of nested models. For example, comparison of a model in which no constraints are imposed (model 1) with one in which equality constraints are specified for all lower order factor loadings (model 2) would constitute a nested model comparison. Based on Cheung and Rensvold (2002) and Chen's (2007) proposal that the difference in CFI and RMSEA values (Δ*CFI and Δ*RMSEA) equal to or less than 0.01 can rightfully serve as viable evidence of invariance and judgement of the overall models exhibiting adequate fit, the hierarchical structure *CFIs and *RMSEA were compared.

Table 23:

Goodness-of-fit and comparative statistics for tests for invariance of BDI–II hierarchical structure across time

	Model and constraints	*S-B* χ^2	*df*	**CFI*	*SRMR*	**RMSEA*	*90% *RMSEA CI*	*Model comparison*	*Δ*CFI*	*Δ*RMSEA*
1.	Configural invariance	691.20	390	0.89	0.06	0.03	0.03, 0.04			
2.	Lower-order factor loadings invariant	688.31	393	0.89	0.06	0.04	0.03, 0.04	2 vs. 1	0.00	0.01
3.	Factor loadings and error variances	649.24	389	0.91	0.06	0.03	0.03, 0.04	3 vs. 1	0.02	0.00

Note. **p* < .001; df = degrees of freedom; S-B χ^2 = Satorra-Bentler scaled Chi-square test; *CFI = robust Comparative Fit Index; *RMSEA = robust root mean square error of approximation and its 90% confidence interval; SRMR = standardized root mean-square residual; Δ*CFI = Comparative Fit Index difference value; Δ*RMSEA = robust root mean square error of approximation difference value.

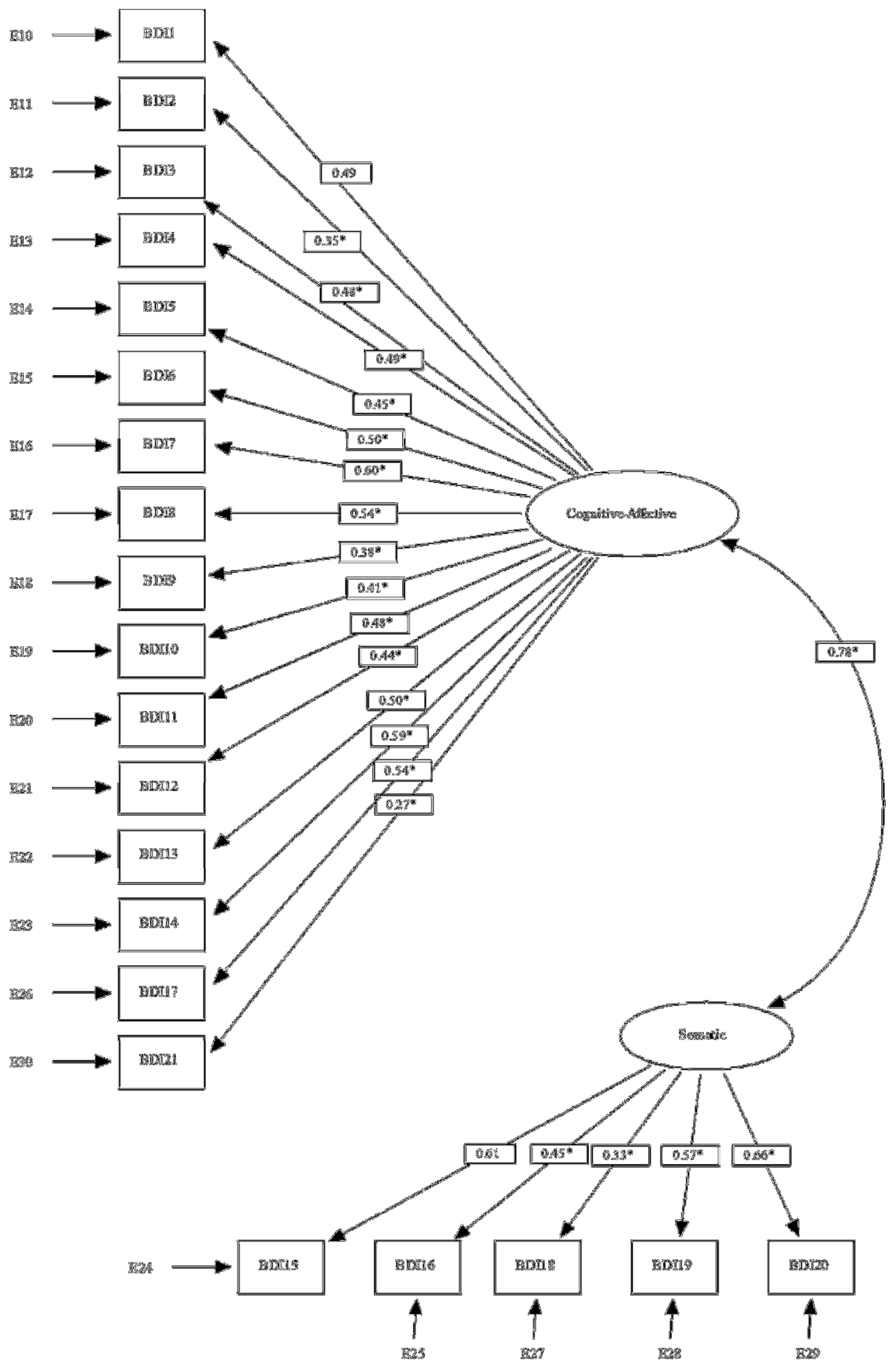
Accordingly, given an overall fit of 0.89 and no change in the overall fit between models 1 and 2 (Δ*CFI = 0.00 and Δ*RMSEA = 0.01), it can be concluded that the lower order factor loadings were not invariant across time. Although, in a final, extremely stringent evaluation of invariance, the equality of measurement error variances across time, the goodness of fit increased a notch to 0.91 (model 3), results still yielded a Δ*CFI of 0.02 and Δ*RMSEA of 0.01, rejecting invariance across time.

4.2.8 *Stage 7 analyses: Examination of Beck et al.'s (1996) two-factor model*

This SEM analysis tested the hypothesized model by Beck et al. (1996) (i.e., the original two-factor model). For the structural model, the significant χ^2 value and its ratio to df is greater than 1.5, but with adequate fit indices, B-B NNFI = 0.90, CFI = 0.91, and acceptable SRMR = 0.04 and RMSEA = 0.04 with 90% CI for RMSEA of 0.03 to 0.04, that agree with the suggested cut-off values in the literature. This implied that the hypothesized two-factor model fits the data well (S-B χ^2 = 393.27, df = 188, $p < 0.0001$, χ^2/df=2.09). The final model for the study is presented in figure 3.

Figure 3:

Cognitive-affective/somatic model (Beck et al., 1996)



Note. The values on the path diagram are standardized regression coefficients, with arrows pointing from latent variables to the observed variables.

The parameter estimates of the *Beck et al.'s (1996)* SEM analysis (figure 3) are shown in table 24. All path coefficient estimates have the expected signs. The magnitudes of the standardized path coefficient estimates, for the measurement components of the model, suggest that the items BDI-1, BDI-3, BDI-4, BDI-5, BDI-6, BDI-7, BDI-8, BDI-10, BDI-11, BDI-12, BDI-13, BDI-14, and BDI-17 have a stronger effect on Factor 1: “Cognitive-Affective” than BDI-2, BDI-9, and BDI-21, while BDI-15, BDI-19 and BDI-20 have a stronger effect on Factor 2: “Somatic” than BDI-16 and BDI-18.

Table 24:

Maximum likelihood estimates of measurement and structural path coefficients for the CFA data (n = 919)

Items	*Coefficient (Robust SE)*	*Robust t–ratio**	*Standardized Coefficient*
F1			
BDI-1	1.00		0.49
BDI-2	0.60(0.09)	6.69	0.35
BDI-3	0.99(0.13)	7.93	0.48
BDI-4	1.19(0.12)	9.84	0.49
BDI-5	0.91(0.10)	8.86	0.45
BDI-6	1.44(0.15)	9.80	0.50
BDI-7	1.36(0.13)	10.77	0.60
BDI-8	1.68(0.17)	10.01	0.54
BDI-9	0.48(0.08)	6.13	0.38
BDI-10	1.48(0.17)	8.84	0.41
BDI-11	1.27(0.13)	9.91	0.48
BDI12	1.16(0.12)	9.57	0.44
BDI-13	1.36(0.14)	9.41	0.50
BDI-14	1.25(0.12)	10.65	0.59
BDI-17	1.32(0.12)	11.05	0.54
BDI-21	0.81(0.14)	6.02	0.27
F2			
BDI-15	1.00		0.61
BDI-16	0.94(0.09)	10.43	0.45
BDI-18	0.55(0.08)	7.41	0.54
BDI-19	1.14(0.10)	11.36	0.57
BDI-20	1.17(0.09)	13.10	0.66

Note. All path coefficients are significant at 5% level of significance.

With respect to the latent factors components of the model (figure 3), the results show that the two factors have statistically significant associations between them (i.e., Cognitive-Affective and Somatic have r = 0.78 with p-value < 0.001).

4.3 Conclusion

This section reported findings of the three-factor structure of the BDI-II among South African collegiate students. This three-factor structure was also found to be invariant across time, race and gender groups.

CHAPTER 5
DISCUSSION, RECOMMENDATIONS AND LIMITATIONS OF THE STUDY

5. Introduction

This section discusses the results and examines whether they are consistent or not with the findings of previous research. The chapter also concludes with recommendations and limitations of the current study.

5.1 Discussion

The present study endeavoured to validate the use of the BDI–II with South African university students. To this end, the researcher identified, tested, and cross-validated the factor structure of the BDI–II for two independent samples of randomly split data using EFA and CFA. Consistent with previous studies of the BDI for nonclinical university students (e.g., Buckley et al., 2001; Byrne et al., 1998; Byrne et al., 1993; Byrne et al., 1995; Byrne et al., 2004; Carmody, 2005; Osman et al., 1997; Osman et al., 2008; Vanheule et al., 2008; Wu & Chang, 2008), our findings revealed that the data for South African university students is best represented by a three-factor model comprising of Negative attitude, Performance difficulty, and Somatic complaints (see figure 1). This model was adequately well-fitting, with no evidence of possible misspecification.

With the evidence of a robust and parsimonious factor structure, we proceeded next to test for its internal consistency reliability and for its stability across two time points. Naturally, only the total score internal consistency is reported, and in this regard, findings from the current study were consistent with those reported for the BDI–II (Beck et al., 1996; Byrne et al., 2004; Dozois et al., 1998). For the sake of comprehensiveness, internal consistency related to each of the three factors (Negative attitude, Performance difficulty, and Somatic complaints) was also assessed. In this instance, all values were adequate, with the weakest findings being associated with the Somatic complaints factor. However, the fact that somatic symptoms tapped by the 6 items comprising this scale are rarely found in the normal population likely accounts for its fairly low reliability.

Similarly, we considered overall depression score stability to be quite acceptable, particularly that there is a possibility of extraneous factors such as life events and stress intervening over the 2 weeks interval between the test-retest administration of the BDI-II. The stability of the BDI-II across time is consistent with past findings (e.g., Aasen, 2001; Beck et al., 1996; Byrne et al., 2004; Dozois et al., 1998). Still, some stability would nonetheless be expected, as individuals are likely to have enduring trait-like tendencies in their mood responsiveness (Byrne et al., 2004). For instance, Tems and colleagues reported that, even following remission of their disorder, adolescents who have been hospitalised reported higher levels of depressive symptoms than did their non-hospitalised controls; albeit, these levels were lower than when they were initially hospitalised (Tems, Stewart, Skinner, Hughes, & Emslie, 1993).

It was also considered prudent to examine correlations between depression and theoretically linked external criteria. Based on theory and empirical research, scores on depression should correlate with risk factors as well as with environmental precipitants (Byrne et al., 2004). As indicated earlier, correlations between depression and hopelessness, and perceived stress would be expected to be high. Inversely, its correlations with self-efficacy and self-esteem would be expected to be lower. True to prediction, depression correlated most strongly with hopelessness and perceived stress, and less so, albeit still significantly, with self-esteem. These findings were consistent with those of earlier studies that demonstrated congruity between the correlates of depressed mood reported in Western and Asian studies (Byrne et al., 2004; Stewart, Betson, Lam, Chung, Ho, & Chung, 1999). Cross-sectional and longitudinal studies suggest that low self-esteem prospectively predicts depression (e.g., Evraire & Dozois, 2011; Hammen, 2005; Joiner, 2000; Morley & Moran, 2011; Kernis et al., 1998; McPherson & Lakey, 1993; O'Brien et al., 2006; Orth et al., 2009; Orth et al., 2008; Roberts, 2006; Roberts & Monroe, 1992).

Contrary to Western literature in support of self-efficacy as an important protective variable for depression (Bandura, 1997), it was found to be correlated strongly and positively with depression in the present study. This finding is consistent with cross-cultural theory that suggests that, in collective cultures (e.g., African), beliefs that emphasize internal sense of personal worth, efficacy, and control may be less significant than in individualistic cultures (Markus & Kitayama, 1994). True to expectation, depression as measured by the BDI-II strongly and positively correlated with depression as measured by HSCL-15 providing evidence for convergent validity. As discussed elsewhere, there are contradictions in the literature on the relationship between depression and anxiety. However, the present study found a strong and positive correlation between depression and anxiety. This is contrary to the cognitive model of depression and anxiety, which says that the two disorders are distinguishable by their cognitive content, and thus should be negatively correlated (Beck, 1967; Beck et al, 1979). This result is consistent with the considerable evidence on the comorbidity of the two disorders (Joiner, 1996; Maser & Cloninger, 1990).

With respect to factorial invariance, evidence of MI in the context of the three-factor structure of the BDI-II across race, gender and across a 2-week time lag was established. MI was established at the level of configural, metric and scalar invariance. Specifically, across models in which there was increasingly restricted parameterization on the variance/covariance matrices of the indicators, there was consistent evidence that the three-factor structure provided robust fit with the data. Furthermore, the Δ*CFI and Δ*RMSEA values for comparisons between models 1 (configural model), 2, 3 and 4 were all negligible. However, there was evidence of DIF characterized by non-invariant items for both race and gender in the test for MI. Racial differences were identified for three BDI-II items [item 2 (pessimism); item 5 (guilt); item 19 (concentration)], while gender differences were also identified for three BDI-II items [Item 11 (agitation); Item 14 (worthlessness); Item 15 (loss of energy)].

Although the existence of racial and gender differences in depressive symptoms is well established, the direction and reasons for the differences are equivocal (Nolen-Hoeksema, Larson, & Grayson, 1999). A variety of social, cultural and personality explanations for greater vulnerability of certain races and gender to specific depressive symptoms have been offered. For instance, level of Westernization often has been associated with variations in the manifestation of depression. People from Western cultures are said to psychologize their depression (i.e., report emotional and cognitive report of distress), whereas people from non-Western cultures are said to somatize their depression (i.e., report distress in the form of bodily complaints and physiological symptoms) (Canino, Rubio-Stipec, Canino, & Escobar, 1992; Katon, Kleinman, & Rosen, 1982; Marsella 1980).

A more psychological report of depression has been found among more Westernized groups of non-Western societies (Kwang-Iel, Dongen, & Dae-Ho, 1999). Somatization of depressive symptoms has been observed in various cultures, such as in Africa (Abiodun, 1995), China (Kwang-Iel et al., 1999), United Arab Emirates (Hamdi, Amin, & Abou-Saleh, 1997), Iraq (Bazzoui, 1970), and India (Teja, Narang, & Aggarwal, 1971). Therefore, that the endorsement of pessimism (Item 2) should be higher for blacks than whites is inconsistent with other depression research in reporting the tendency of blacks to somatize than psychologize.

However, this finding is consistent with research reporting no significant demographic differences between people who tend to somatize and those who psychologize (Blazer, Landerman, Hays, Simonsick, & Saunders, 1998; Razali & Hasanah, 1999). A possible explanation for this unexpected direction in difference can be the fact that South Africa is more westernized/acculturated and therefore we expected that evaluations of self by blacks, as measured by the BDI-II, would necessarily reflect the worldview espoused by the society. This finding, of higher pessimism in blacks, is also contradictory to the implication of external locus-of-control beliefs in blacks in the explanation of depression symptoms across cultures. An external locus-of-control in blacks is attributed to the fact that blacks have limited opportunities as an economic minority group in South Africa. So, theoretically, blacks are expected to report more feelings of being punished and less self-blame. That the endorsement of a somatic item (Item 5) should be higher among blacks than whites is consistent with other depression research, which reports that non-Western individuals tend to somatize their depression than the westernized, usually white, individuals (Canino et al., 1992; Katon et al., 1982; Marsella, 1980). As such, blacks typically show greater concern for somatic symptoms than whites. This is consistent with the suggestion of Jenkins, Kleinman and Good (1991) that somatization may serve as a coping style that "protects" the depressed individual from feelings of self-blame and hopelessness.

Kleinman (1988) explains that somatization as an 'idiom of distress' is more common in cultures where stigma is connected with psychiatric problems and the expression of emotional distress is inhibited (i.e., black society in South Africa). Similarly, the stronger endorsement of Item 19 (concentration) by whites than by blacks is also consonant with the literature, in that whites typically psychologize their depression. This may also be attributed to the fact that people in individualistic cultures may be more in tune with their private emotional states, while those socialized in collectivistic cultures are more responsive to the promotion of the welfare of their in-group and thus emotions are used more strategically (Canino et al., 1992). Accordingly, emotional states are a more immediate and prominent source for the self-efficacy appraisals of individuals raised in idiocentric systems than in allocentric cultures.

Likewise, the reasons and directions of gender differences in depressive symptomatology are also not yet well understood. One model suggests that females are more prone to exhibit a cognitive style characterized by negative self-evaluation and rumination/ruminative coping, which in turn may predispose them to depression (Garber & Martin, 2002; Gilligan & Attanucci, 1988; Hankin & Abramson, 2001; Nolen-Hoeksema et al., 1999; Siegel et al., 1999). The presence of gender-based differences in cognitive symptoms is documented among both psychiatric patients and normal controls. Based on these findings, females have typically shown greater vulnerability to negative self-evaluation, more mood symptoms and self-deprecation than males (Nolen-Hoeksema et al., 1999). Feminist theories of women's greater vulnerability to depressive symptoms compared with that of men generally attribute this susceptibility to the negative consequences of women's lower social status and power (Nolen-Hoeksema et al., 1999). Because of this lower status and power, women experience more negative events and have less control over important areas in their lives than men. This assertion is true for phallocentric societies. Thus, it is not surprising that Item 11 and 14, measuring agitation and worthlessness, received significantly greater endorsement by females than males. Indeed, the impact of gender-related stereotypes on the expression of emotional distress is now widely known (see Nolen-Hoeksema, 1987; Nolen-Hoeksema & Girgus, 1994). Typically, females are inclined to openly acknowledge and express emotional weaknesses or negative affect, whereas males tend to deny these feelings.

That the endorsement of an item measuring loss of energy (Item 15) should be higher for females than males is an anomaly, since depression research describes males as more behaviorally-oriented than females (Baron & Joly, 1988; Nolen-Hoeksema & Grigus, 1994; Vredenburg et al., 1986). Males are expected to show a characteristically greater concern for their ability to perform than females. Thus, the fact that this item received higher endorsement by females in the present study must be regarded as somewhat atypical. However, one possible explanation for this gender difference with respect to South Africans may lie with the post-1994 change in gender roles and social order. There is a trend towards the sharing of responsibilities and societal expectations between males and females.

In addition, scalar invariance was established with two noninvariant item intercepts for race (items 5 and 14) and also for gender with three noninvariant item intercepts (items 11, 14 and 18), providing empirical evidence of construct validity for the BDI-II for purposes of making race and gender latent mean comparisons. This findings of two noninvariant intercepts across race and three intercepts across gender implied that there is differential ARS bias (Cheung & Rensvold, 2000) for the BDI-II across these groups. That is, blacks systematically endorse a relatively higher item response in Item 5, whereas whites systematically endorse a higher item response in Item 14. Likewise, males systematically endorse higher item responses in Item 18 than females, whereas females systematically record higher scores in Items 11 and 14. The presence of differential ARS of the BDI-II in the current study confirms Wu's (2010a, b) findings.

The evidence of scalar invariance also implies that clinicians and researchers can use the BDI-II with greater confidence given the generalizability of the instrument's properties between blacks and whites, and male and female university students in South Africa. Contrary to most previous studies investigating race and gender differences on overall depression (e.g., Hankin, Abramson, Moffitt, Silva, McGee, & Angell, 1998; Schuch, Roest, Nolen, Penninx, & de Jonge, 2014), the present study examined group differences at specific factor levels. Results revealed significant latent mean and observed mean differences that favored blacks in terms of Performance difficulty and Somatic complaints factors. As for gender, latent mean and observed mean differences were significant in terms of Negative attitude and Somatic complaints and favored females.

The results on differential race endorsements of items are explainable if we consider that people from non-Western cultures tend to somatize their depression (i.e., report affect in the form of bodily complaints and physiological symptoms) (Canino et al., 1992; Katon et al., 1982; Marsella 1980). Likewise, results on gender are explicable if one considers that most symptoms associated with the Negative attitude factor are cognitive in nature. Because women are more apt to exhibit a cognitive symptom pattern characterized by negative self-evaluation, this may prejudice them to be depressed on cognitively inclined depression instruments such as the BDI-II (e.g., Hankin & Abramson, 2001; Nolen-Hoeksema, Larson, & Grayson, 1999).

The current study also demonstrated the advantages of using latent mean analyses as opposed to observed mean differences for understanding race and gender differences. For example, table 14 showed that females have significantly higher observed scores on the Performance difficulty factor than males even when conducting one-tailed hypothesis testing. Thus, one may conclude that there were gender differences on this factor based on observed mean differences. However, the latent mean analyses (see table 14) revealed that females did not have significantly higher endorsements on the Performance difficulty factor than did males when conducting one-tailed hypothesis testing. As noted elsewhere, observed mean differences are not identical to latent mean differences. The former confounds factor and indicator variance, and the latter is calculated by partialing out variance attributable to measurement error (Cheung & Rensvold, 2000). The degree of confounding may have critical effects on observed mean differences. Therefore, when evaluating race and gender differences on depression, one must be cautious in making final conclusions merely based on observed mean scores. Rather, both latent and observed mean scores should be taken into account.

The present study found two noninvariant intercepts across race (item 5 and 14) and three intercepts across gender (item 11, 14 and 18) in the BDI-II with negligible effects on latent mean differences. Although scalar invariance properties of the BDI-II across race and gender groups is established, researchers and practitioners are urged to interpret race and gender differences corresponding to these noninvariant item intercepts with caution. These intercepts failed to display equivalent measurement across race and gender groups, signifying that these items overestimated the corresponding factor for one group. In that regard, the item depicting guilt overestimated its respective factors for blacks, whereas that depicting worthlessness overestimated its factor for whites. Similarly, items measuring agitation and worthlessness overestimated their respective factors for females, whereas that measuring appetite overestimated its respective factor for males.

Collectively, these results provide strong evidence for MI for the BDI-II across race, gender and time, leading to the conclusion that the BDI-II does not measure different hypothetical traits for one group (race and gender) than another, or across time. These findings suggest that blacks and whites and males and females: (a) have the same structure of the BDI-II (e.g., the same number of factors and each factor is associated with the same items) and (b) have equal strengths of relations between the underlying construct and specific scale items. Since factorial invariance was obtained for analyses constraining factor structure and loadings, these results suggest that it is also appropriate to compare correlates of depressive symptoms across groups (Whisman et al., 2012). These findings corroborate past research on the MI of the BDI-II within student populations across race, gender and time (Byrne et., 1993, 1994; Byrne et al., 1996; Byrne et al., 1996; Byrne et al., 2007; Campbell et al., 2009; Hooper et al., 2012; Whisman et al., 2012; Wu, 2010; Wu & Huang, 2012).

The present study also evaluated the fit of established factor structure models proposed by Byrne et al. (2004) and Beck et al. (1996) for our sample, and compared them with our hypothesized model. Both models (i.e., Byrne et al.'s three-factor and Beck et al.'s two-factor models) demonstrated good fits with our data, although comparatively inferior to our hypothesized three-factor model. For instance, Byrne et al.'s model had to be modified and improved (e.g., inclusion of 4 correlations between measurement errors) in order to obtain a satisfactory fit. Therefore, our hypothesized three-factor model has superior fit for the present student data than Beck et al.'s (1996) original two-factor model and Byrne et al.'s (2004) three-factor model. Furthermore, Byrne et al.'s (2004) hypothesized model was found to be invariant across race (black and white) and gender groups, but longitudinal invariance (across a two-week time lag) with this student sample was rejected.

It can therefore be concluded that for the BDI-II, black and white South African students' responses demonstrated relatively little difference in the relationship between their responses and the relationship to the latent variable. The BDI-II appears to measure dysphoria roughly equivalently across black and white, and male and female student respondents in South Africa.

5.2 Recommendations

Research suggests that if MI has not been established, valid group comparisons cannot be made (e.g., Horn & McArdle, 1992; Little, 1997). In the current study, the evidence of MI was established, indicating that clinicians and researchers can use the BDI-II with confidence given the generalizability of the instrument's properties between black and white, male and female university students and stability across time. Moreover, comparisons of latent mean differences of depression between black and white, and male and female university students are possible.

Comparisons of latent mean differences are necessary because assessing specific factor differences is more meaningful than assessing group differences on BDI-II overall scores. Especially when there is vast research about which particular factor tends to favor blacks or whites, or males or females, exploring and adjusting how that particular factor is weighted to form the depression is cogently suggested. This is necessary because it is possible that some factors have a higher criterion-related validity than others (e.g., cognitive factor with depression in women).

Researchers and clinicians are advised to take both latent factor scores and observed factor scores into account, so as to develop a better appreciation of gender differences and circumvent risks associated with measurement error. The awareness of these possible differences may be helpful to clinicians during treatment planning. For instance, if worthlessness and body image discontent are more prominent issues in depression for females than males, then cognitive interventions aimed at alleviating worthlessness and improving body image may particularly be helpful for some depressed females.

Our findings reveal that the data for South African students is best represented by a hierarchically structured model defined by three lower-order factors comprising Negative Attitude (BDI-1, BDI-2, BDI-3, BDI-7, BDI-9, BDI-11, BDI-14 and BDI-17), Performance Difficulty (BDI-12, BDI-13, BDI-15, BDI-

16, BDI-18, BDI-19 and BDI-20), and Somatic complaints (BDI-4, BDI-5, BDI-6, BDI-8, BDI-10 and BDI-21) (see figure 1).

5.3 Limitations

While the present results are significant for studies using student samples, it is unclear whether comparable results would be obtained with clinical samples and nonclinical samples of individuals at different age levels either those characteristic of typical students. Additionally, the findings do not address MI of the BDI-II for subgroups (ethnic groups) that exist within larger categories defined by race (black and white), particularly as university students may represent the most acculturated members of their communities. This is due to the fact that race was defined by self-reports and that there is considerable diversity within the broad category of race in South Africa. As such, the use of university students may underestimate the impact of values and practices, therefore, attitudes and behaviour of minority groups on BDI-II responses.

5.4 Conclusion

This chapter discussed and contextualized the results, proffered recommendations for future studies and concluded with limitations inherent in the current study.

REFERENCES

Aasen, H. (2001). *An empirical investigation of depression symptoms: Norms, psychometric characteristics and factor structure of the Beck Depression Inventory-II.* Unpublished MA Dissertation. University of Bergen: Bergen, Norway.

Abiodun, O. A. (1995). A study of somatization among primary care patients in Nigeria. *International Journal of Mental Health*, *23*, 81-89.

Abrahams, F. (2002). The (un)fair usage of the 16PF (SA92) in South Africa: A response to C. H. Prinsloo and I. Ebersöhn. *South African Journal of Psychology*, *32,* 58-61.

Abrahams, F., & Mauer, K. F. (1999). Qualitative and statistical impact of home language on responses to the items of the Sixteen Personality Factor Questionnaire (16PF) in South Africa. *South African Journal of Psychology*, *29*, 76-86.

Aiken, L. S., Stein, J. A., & Bentler, P. M. (1994). Structural equation analyses of clinical subpopulation differences and comparative treatment outcomes: Characterizing the daily lives of drug addicts. *Journal of Consulting & Clinical Psychology, 62,* 488–499.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th ed.).* Washington, DC: Author.

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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.).* Arlington, VA: American Psychiatric Publishing.

Anastasi, A. (1986). Evolving concepts of test validation. *Annual Review of Psychology*, *37*, 1-15.

Allan, A. (2001). *The law for psychotherapists and counsellors*. Somerset West, South Africa: Inter-Ed.

Al-Musawi, N. M. (2001). Psychometric properties of the Beck Depression Inventory-II with university students in Bahrain. *Journal of Personality Assessment*, *77*, 568-579.

Al-Turkait, F. A., & Ohaeri, J. U. (2010). Dimensional and hierarchical models of depression using the Beck Depression Inventory-II in an Arab college student sample. *BMC Psychiatry*, *10*(60), 1471-2440.

Al-Turkait, F. A., & Ohaeri, J. U. (2012). Exploratory factor analytical study of depressive symptomatology: An Arab experience with a sample comprising college students, using the revised Beck Depression Inventory. *Annals of Saudi Medicine*, *32(1),* 19-26.

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

Azocar, F., Arean, P., Miranda, J., & Munoz, R. F. (2001). Differential item functioning in a Spanish translation of the Beck Depression Inventory. *Journal of Clinical Psychology*, *57(3),* 355-365.

Baldwin, D. R., Harris, S. M., & Chambliss, L. N. (1997). Stress and illness in adolescence: Issues of race and gender. *Adolescence*, *32*, 839-853.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Baron, P., & Joly, E. (1988). Sex differences in the expression of depression in adolescents. *Sex Roles, 18,* 1 –7.

Bazzoui, W. (1970). Affective disorders in Iraq. *British Journal of Psychiatry*, *117,* 195-203.

Beck, A. T. (1967). *Depression: Clinical, experimental and theoretical aspects*. Philadelphia, PA: University of Pennsylvania.

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

Beck, A. T., & Steer, R. A. (1987). *Beck Depression Inventory: Manual*. San Antonio, TX: The Psychiatric Corporation.

Beck, A. T., Steer, R. A., & Brown, G. K. (2000). *BDI-FastScreen for medical patients*. San Antonio, TX: Pearson.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II Manual (2nd ed.)*. San Antonio: Harcourt Brace & Company.

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

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and –II in psychiatric outpatients. *Journal of Personality Assessment*, *67(3),* 588-597.

Beck, A. T., Steer, R. A., Epstein, N. & Brown, G. (1990). The Beck self-concept test. *A Journal of Consulting & Clinical Psychology*, *2,* 191-197.

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., Moch, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561-571.

Beck, A. T., Weissman, A., Lester, D. & Trexler, L. (1974). The measurement of pessimism: The hopelessness scale. *Journal of Consulting & Clinical Psychology*, *42,* 861–865.

Bentler, P. M. (2005). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software, Inc.

Bentler, P. M. (2004). *EQS 6.1: Structural equations program manual*. Encino, CA: Multivariate Software, Inc.

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

Bentler, P. M. (1989). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bernal, G., Bonilla, J., & Santiago, I. (1995). Confiabilidad y validez de construcción lógica de dos instrumentos para medir sintomatología psicológica en una muestra clínica: El Inventario de Depresión de Beck y la Lista de Cotejo de Síntomas-36. *Revista Latinoamericana de Psicología*, *27,* 207–229. Abstract retrieved from http://prhsj.rcm.upr.edu/index.php/prhsj/article/viewFile/315/206

Blascovich, J., & Tomaka, J. (1991). Measures of self-esteem. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes.* Vol. 1. (pp. 115-160). San Diego, CA: Academic.

Blazer, D. G., Landerman, L. R., Hays, J. C., Simonsick, E. M., & Saunders, W. B. (1998). Symptoms of depression among community-dwelling elderly African Americans and White older adults. *Psychological Medicine*, *28*, 1311-1320.

Bonilla, J., Bernal, G., Santos, A., & Santos, D. (2004). A revised Spanish version of the Beck Depression Inventory: Psychometric properties with a Puerto Rican sample of college students. *Journal of Clinical Psychology*, *60,* 119-130.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York, NY: Wiley.

Bos, S. C., Pereira, A. T., Marques, M., Maia, B., Soares, M. J., Valente. J., et al. (2008). The BDI-II factor structure in pregnancy and postpartum: Two or three factors? *European Psychiatry, 10*, 1-7.

Bouchard, T. J. (1976). Unobtrusive measures: An inventory of uses. *Sociological Methods & Research*, *4,* 267-300.

Boughton, S., & Street, H. (2007). Integrated review of the social and psychological gender differences in depression. *Australian Psychologist*, *42,* 187-197.

Bracken, B., & Reintjes, C. (2010). Age, race, and gender differences in depressive symptoms: A lifespan developmental investigation. *Journal of Psychoeducational Assessment, 28(1),* 40–53.

Brantley, P. J., Dutton, G. R., & Wood, K. B. (2004). The Beck Depression Inventory—II (BDI–II) and the Beck Depression Inventory—Primary Care (BDI–PC). In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and assessment outcome (3rd ed.)*(pp. 313–326). Mahwah, NJ: Erlbaum.

Brennan, R. L. (2004). *Revolutions and evolutions in current educational testing (CASMA Research Rep. No. 4)*. Iowa City: University of Iowa.

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(3),* 197- 204.

Butler, A. C., Hokanson, J. E., & Flynn, H. A. (1994). A comparison of self-esteem lability and low trait self-esteem as vulnerability factors for depression. *Journal of Personality & Social Psychology*, *66,* 166–177.

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming (2nd ed.).* Mahwah, NJ: Erlbaum.

Byrne, B. M., & Baron, P. (1993). The Beck Depression Inventory: Testing and cross-validating an hierarchical structure for nonclinical adolescents. *Measurement & Evaluation in Counseling & Development*, *26,* 164–178.

Byrne, B. M., & Baron, P. (1994). Measuring adolescent depression: Tests of equivalent factorial structure for English and French versions of the Beck Depression Inventory. *Applied Psychology*: *An International Review*, *43,* 33–47.

Byrne, B. M., & Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-Cultural Psychology*, *30,* 557–576.

Byrne, B. M., Baron, P., & Balev, J. (1998). The Beck Depression Inventory: A cross-validated test of factorial structure for Bulgarian adolescents. *Educational & Psychological Measurement*, *58,* 241–251.

Byrne, B. M., Baron, P., & Balev, J. (1996). The Beck Depression Inventory: Testing for its factorial validity and invariance across gender for Bulgarian adolescents. *Personality & Individual Differences*, *21,* 641–651.

Byrne, B. M., Baron, P., & Campbell, T. L. (1994). The Beck Depression Inventory (French version): Testing for gender-invariant factorial structure for nonclinical adolescents. *Journal of Adolescent Research*, *9,* 166–179.

Byrne, B. M., Baron, P., & Campbell, T. L. (1993). Measuring adolescent depression: Factorial validity and invariance of the Beck Depression Inventory across gender. *Journal of Research on Adolescence*, *3,* 127–143.

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

Byrne, B. M., Baron, P., Larsson, B., & Melin, L. (1996). Measuring depression for Swedish nonclinical adolescents: Factorial validity and equivalence of the Beck Depression Inventory across gender. *Scandinavian Journal of Psychology*, *37*, 37–45.

Byrne, B. M., & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, *13,* 287–321.

Byrne, B. M., Stewart, S. W., Kennard, B. D., & Lee, P. W. H. (2007). The Beck Depression Inventory II: Testing for measurement equivalence and factor mean differences across Hong Kong and American adolescents. *International Journal of Testing*, *7,* 293-309.

Byrne, B. M., Stewart, S. M., & Lee, P. W. H. (2004). Validating the Beck Depression Inventory–II for Hong Kong community adolescents. *International Journal of Testing*, *4,* 199–216.

Camara, W., Nathan, J., & Puente, A. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research & Practice*, *31(2),* 141-154.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin*, *56,* 1-105.

Campbell, M. H., Roberti, J. W., Maynard, D., & Emmanuel, M. K. (2009). Factorial structure of depressive symptoms in Anglophone Caribbean university students: Psychometric properties of the Beck Depression Inventory-II. *InterAmerican Journal of Psychology*, *43(3),* 577-585.

Canino, I. A., Rubio-Stipec, M., Canino, G., & Escobar, J. I. (1992). Functional somatic symptoms: A cross-ethnic comparison. *American Journal of Orthopsychiatry*, *62,* 605-612.

Carmody, D. P. (2005). Psychometric characteristics of the Beck Depression Inventory-II with college students of diverse ethnicity. *International Journal of Psychiatry in Clinical Practice*, *9*, 22-28.

Cavanaugh, S. (1983). The prevalence of emotional and cognitive dysfunction in a general medical population: Using the MMSE, GHQ and BDI. *General Hospital Psychiatry, 5,* 15-24.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464-504.

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new General Self-Efficacy Scale. *Organizational Research Methods*, *4(1),* 62-68.

Chen, G., Gully, S. M., Whiteman, J. A., & Kilcullen, B. N. (2000). Examination of relationships among trait-like individual differences, state-like individual differences, and learning performance. *Journal of Applied Psychology*, *85(6),* 834-847.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling*, *12*, 471-492.

Cheung, F. M., Leong, F. T. L., & Ben-Porath, Y. S. (2003). Psychological assessment in Asia: Introduction to the special section. *Psychological Assessment, 15*, 243–247.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9,* 233–255.

Cheung, P. C., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equation modeling. *Journal of Cross-Cultural Psychology*, *31,* 187-212.

Claassen, N. C. W. (1997). Culture differences, politics and test bias in South Africa. *European Review of Applied Psychology*, *47*, 297-307.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health & Social Behavior*, *24,* 385-396.

Cole, J. C., Grossman, I., Prilliman, C., & Hunsaker, E. (2003). Multimethod validation of the Beck Depression Inventory-II and Grossman-Cole Depression Inventory with an inpatient sample. *Psychological Reports*, *93,* 1115-1129.

Conoley, C. W. (1992). Review of the Beck Depression Inventory [Revised Edition]. In J. J. Kramer & J. C. Conoley (Eds.), *Eleventh mental measurements yearbook* (pp. 78-79). Lincoln, NE: Buros Institute of Mental Measurements.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, *10(7),* 1-9.

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

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Belmont, CA: Wadsworth.

Crocker, L. (1997). The great validity debate. *Educational Measurement: Issues & Practice*, *16(2),* 4.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52,* 281-302.

Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed.). New York, NY: Harper & Row.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer (Ed.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum Associates.

De Bruin, G. P., Swartz, L., Tomlinson, M., Cooper, P.J., & Molteno, C. (2004). The factor structure of the Edinburgh Postnatal Depression scale in a South African peri-urban settlement. *South African Journal of Psychology*, *34,* 113-121.

de Jonge, P., Ormel, J., van den Brink, R. H., van Melle, J. P., Spijkerman, T. A., Kuijper, A., et al. (2006). Symptom dimensions of depression following myocardial infarction and their relationship with somatic health status and cardiovascular prognosis. *American Journal of Psychiatry*, *163(1),* 138-144.

De Klerk, J. J., Boshoff, A. B., & Van Wyk, R. (2009). Measuring meaning in life in South Africa: Validation of an instrument developed in the USA. *South African Journal of Psychology, 39(3),* 314-325.

DeLuca, C., Klinger, D. A., Searle, M., & Shulha, L. M. (2010). Developing a curriculum for assessment education. *Assessment Matters*, *2,* 20-42.

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, I. (1974). The Hopkins Symptoms Checklist (HSCL): A measure of primary symptom dimensions. *Modern Problems of Pharmacopsychiatry*, *7*, 79-110.

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

Drennan, G. (1988). An investigation into the process of translating meaning across cultures: The Beck Depression Inventory. *Unpublished Psychology Honours research project*, University of Cape Town, Cape Town.

Dunkel, D., Froehlich, S., Antretter, E., & Haring, C. (2002). Replication of a two-factor model of the Beck Depression Inventory in alcohol dependents and suicide attempters. *Psychopathology, 35(4),* 228-233.

Dutton, G. R., Grothe, K. B., Jones, G. N., Whitehead, D., Kendra, K., & Brantley, P. J. (2004). Use of the Beck Depression Inventory—II with African American primary care patients. *General Hospital Psychiatry*, *26,* 437–442.

Eaton, N. R., Keyes, K. M., Krueger, R. F., Balsis, S., Skodol, A. E., Markon, K. E., et al. (2011). An invariant dimensional liability model of gender differences in mental disorder prevalence: Evidence from a national sample. *Journal of Abnormal Psychology, 121(1),* 282-288.

Evraire, L. E., & Dozois, D. J. A. (2011). An integrative model of excessive reassurance seeking and negative feedback seeking in the development and maintenance of depression. *Clinical Psychology Review*, *31*, 1291-1303.

Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling*, *12*, 343–367.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, research version, nonpatient edition* (SCID-I/NP). New York, NY: Biometrics Research, New York State Psychiatric Institute.

Forintos, D. P., Sallai, J., & Rózsa, S. (2010). Adaptation of the Beck Hopelessness Scale in Hungary. *Psychological Topics*, *19(2),* 307-321.

French, B. F., & Finch, W. H. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural Equation Modeling*, *13(3)*, 378-402.

Gagne, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, *41(1),* 65-83.

Garber, J., & Martin, N. C. (2002). Negative cognitions in offspring of depressed parents: mechanisms of risk. In S. H. Goodman, I. H. Gotlib. (Eds.), *Children of depressed parents: Mechanisms of risk and implications for treatment (pp. 121–153)*. Washington: DC American Psychological Association.

Gary, F. A., & Yarandi, H. N. (2004). Depression among southern rural African American women: A factor analysis of the Beck Depression Inventory-II. *Nursing Research*, *53(4),* 251-259.

Ghassemzadeh, H., Mojtabai, R., Karamghadiri, N., & Ebrahimkhani, N. (2005). Psychometric properties of a Persian-language version of the Beck Depression Inventory—Second edition: BDI-II-PERSIAN. *Depression & Anxiety*, *21,* 185-92.

Gladstone, T. R., & Koenig, L. J. (1994). Sex differences in depression across the high school to college transition. *Journal of Youth & Adolescence*, *23,* 643–649.

Gilligan, C., & Attanucci, J. (1988). Two moral orientations. In C. Gilligan., J. V. Ward, J. M. Taylor., B. Bardige. (Eds.), *Mapping the moral domain: A contribution of women's thinking to psychological theory and education*. (pp. 73-86). Cambridge, MA: Harvard University.

Gorsuch, R. L. (1983). *Factor analysis (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum.

Gotlib, I. H., & Hammen, C. L. (1992). *Psychological aspects of depression: Toward a cognitive-interpersonal integration*. New York, NY: John Wiley & Sons.

Gray, D., & Durrheim, K. (2006). The validity and reliability of measures of right-wing authoritarianism in South Africa. *South African Journal of Psychology*, *36,* 500-520.

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(1),* 110-114.

Guion, R. (1977). Content validity: The source of my discontent. *Applied Psychological Measurement*, *1*, 1-10.

Hagborg, W. J. (1993). The Rosenberg Self-esteem Scale and Harter's Self-Perception Profile for adolescents: A concurrent validity study. *Psychology in the Schools*, *30*, 132-136.

Halvorsen, J. O., & Kagee, A. (2010). Predictors of psychological sequelae of torture among South African former political prisoners. *Journal of Interpersonal Violence*, *25*, 989-1005.

Hambrick, J. P., Rodebaugh, T. L., Balsis, S., Woods, C. M., Mendez, J. L., & Heimberg, R. G. (2010). Cross-ethnic measurment equivalence of measures of depression, social anxiety, and worry. *Assessment, 17*, 155-171.

Hamdi, E., Amin, Y., & Abou-Saleh, M. T. (1997). Performance of the Hamilton Depression Rating Scale in depressed patients in the United Arab Emirates. *Acta Psychiatrica Scandinavica, 96*, 416-423.

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

Hammen, C. (2005). Stress and depression. *Annual Review of Clinical Psychology*, *1*, 293–319.

Hankerson, S. H., Fenton, M. C., Geier, T. J., Keyes, K. M., Weissman, M. M., & Hasin, D. S. (2011). Racial differences in symptoms, comorbidity, and treatment for major depressive disorder among black and white adults. *Journal of the National Medical Association*, *103(7),* 576-584.

Hankin, B. L. (2002). Gender differences in depression from childhood through adulthood: A review of course, causes, and treatment. *Primary Psychiatry*, *9*, 32-36.

Hankin, B. L., & Abramson, L. Y. (2001). Development of gender differences in depression: An elaborated cognitive vulnerability transactional stress theory. *Psychological Bulletin*, *127,* 773–796.

Hankin, B. L., Abramson, L. Y., Moffitt, T. E., Silva, P. A., McGee, R., & Angell, K. E. (1998). Development of depression from preadolescence to young adulthood: Emerging gender differences in a 10-year longitudinal study. *Journal of Abnormal Psychology*, *107(1),* 128-140.

Hooper, L. M., Qu, L., Crusto, C., & Huffman, L. E. (2012). Scalar equivalence self-rated depressive symptomatology as measured by the Beck Depression Inventory-II: Do racial and gender differences in college students exist? *Psychology, 3,* 762-774.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18,* 117–144.

Hoyle, R. H. (1991). Evaluating measurement models in clinical research: Covariance structure analysis of latent variable models of self-conception. *Journal of Consulting & Clinical Psychology*, *59,* 67–76.

Human Sciences Research Council. (2007). Beck Depression Inventory II (BDI II). African early child development measures resources. Retrieved from http://wwwhsrc.ac.za/ECD-Measures-125.phtml

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

Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist: Statistics using generalized linear models*. Thousand Oaks, CA: Sage.

James, S., & Prilleltensky, I. (2002). Cultural diversity and mental health: Towards integrative practice. *Clinical Psychology Review*, *22,* 1133–1154.

Jenkins, J. H., Kleinman, A., & Good, B. J. (1991). Cross cultural studies of depression. In J. Becker & A. Kleinman (Eds.), *Psychsocial aspects of depression* (pp. 67-93). Hillsdale, NJ: Lawrence Erlbaum.

Johnson, S. K., DeLuca, J., & Natelson, B. J. (1996). Depression in fatiguing illness: Comparing patients with chronic fatigue syndrome, multiple sclerosis and depression. *Journal of Affective Disorders*, *39*, 21-30.

Johnson, M. E., Neal, D. B., Brems, C., & Fisher, D. G. (2006). Depression as measured by the Beck Depression Inventory-II among injecting drug users. *Assessment, 13*, 168–177.

Joiner, T. E. (1996). A confirmatory factor-analytic investigation of the tripartite model of depression and anxiety in college students. *Cognitive Therapy & Research*, *20*, 521–539.

Joiner, T. E. (2000). A test of the hopelessness theory of depression among youth psychiatric inpatients. *Journal of Abnormal Psychology*, *29,* 167–176.

Kaaya, S. F., Fawzi, M. C., Mbwambo, J. K., Lee, B., Msamanga, G. I., & Fawzi, W. (2002). Validity of the Hopkins Symptom Checklist-25 amongst HIV- positive pregnant women in Tanzania. *Acta Psychiatry Scandinavica*, *106(1),* 9-19.

Kagee, A. (2005). Symptoms of distress and posttraumatic stress among South African former political detainees. *Ethnicity & Health*, *10*, 169-179.

Kagee, A., & Martin. L. (2010). Symptoms of depression and anxiety among a sample of South African patients living with HIV. *AIDS Care*, *22,* 159-165.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38,* 319-342.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: American Council on Education/Praeger.

Katon, W., Kleinman A., & Rosen, G. (1982). Depression and somatization: A review I. *American Journal of Medicine*, *72*, 127-135.

Kernis, M. H., Whisenhunt, C. R., Waschull, S. B., Greenier, K. D., Berry, A. J., Herlocker, C. E., et al. (1998). Multiple facets of self-esteem and their relations to depressive symptoms. *Personality & Social Psychology Bulletin*, *24,* 657-668.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., et al. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the national comorbidity survey. *Archives of General Psychiatry*, *51,* 8-19.

Kim, K. (2005). The relation among fit indexes, power, and sample size in structural equation modeling. *Structural Equation Modeling*, *12*, 368–390.

Kleinman, A. M. (1986). *Social origins of distress and disease: Depression, neurasthenia, and pain in modern China*. New Haven, CT: Yale University.

Kleinman, A. M. (1988). *Rethinking psychiatry: From cultural category to personal experience*. New York, NY: Free.

Kneipp, S. M., Kairalla, J. A., Stacciarini, J., & Pereira, D. (2009). The Beck Depression Inventory II factor structure among low-income women. *Nursing Research*, *58(6),* 400–409.

Koenig, L. J., Isaacs, A. M., & Schwartz, J. A. J. (1994). Sex differences in adolescent depression and loneliness: Why are boys lonelier if girls are more depressed? *Journal of Research in Personality*, *28,* 27–43.

Kojima, M., Furukawa, T. A., Takahashi, H., Kawai, M., Nagay, T., & Tokudome, S. (2002). Cross-cultural validation of the Beck Depression Inventory-II in Japan. *Psychiatric Research, 110,* 291-299.

Kwang-Iel, K., Dongen, L., & Dae-Ho, K. (1999). Depressive symptoms in Koreans, Korean-Chinese and Chinese: A transcultural study. *Transcultural Psychiatry, 36,* 303-316.

Lovibond, S. H. & Lovibond, P. F. (1993). *Manual for the Depression Anxiety Stress Scales (DASS)*. Psychology Foundation Monograph. (Available from The Psychology Foundation, Room 1005 Mathews Building, University of New South Wales, NSW 2052, Australia).

Leff, J. P. (1977). Notions of the emotions. *Mental Health & Society*, *4(5-6),* 308-318.

Leganger, A., Kraft, P., & Røysamb, E. (2000). Perceived self-efficacy in health behaviour research: Conceptualization, measurement and correlates. *Psychology & Health, 15*, 51-69.

Lester, D., & Akande, A. (1995). Depression in Nigerian and American students. *Psychological Reports*, *76*, 906.

Leung, P. W. L., & Wong, M. M. T. (2003). Measures of child and adolescent psychopathology in Asia. *Psychological Assessment*, *15,* 268–279.

Lin, H. C. (2012). *Racial/ethnic differences in depressive symptoms and treatment effect among patients with myocardial infarction from the enhancing recovery in coronary heart disease (ENRICHD) trial*. Unpublished Doctoral Dissertation. University of Miami: Coral Gables, Florida

Linke, S. E., Rutledge, T., Johnson, B. D., Vaccarino, V., Bittner, V., Cornell, C. E., et al. (2009). Depressive symptom dimensions and cardiovascular prognosis among women with suspected myocardial ischemia: A report from the National Heart, Lung, and Blood Institute sponsored Women's Ischemia Syndrome Evaluation. *Archives of General Psychiatry*, *66(5),* 499-507.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: Wiley.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *31*, 53–76.

Loevinger, J. (1957). Objective tests of psychological theory. *Psychological Reports*, *3,* 635-694.

Lu, A. L., Che, H. H., Chang, S. W., & Shen, W. W. (2002). Reliability and validity of the Chinese version of the Beck Depression Inventory II. *Taiwanese Psychiatry*, *16*, 301-310.

Lugo, Y. (1999). *El abuso sexual, los trastornos de ansiedad y la depresión: Estudio de las características psicometricas del Inventario de Ansiedad y la Escala de Depresión de Beck-II en adultos sobrevivientes de abuso sexual* [Disertación].

San Juan (PR): Centro Caribeño de Estudios Postgraduado. Abstract retrieved from http://prhsj.rcm.upr.edu/index.php/prhsj/article/viewFile/315/206

Luszczynska, A., Gutiérrez-Donã, B., & Schwarzer, R. (2005). General self-efficacy in various domains of human functioning: Evidence from five countries. *International Journal of Psychology*, *40(2),* 80-89.

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36(4),* 611-637.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Measurement*, *4(1),* 84-89.

Mahmud, W. M., Awang, A., Herman, I., & Mohamed, M. N. (2004). Analysis of the psychometric properties of the Malay version of Beck Depression Inventory II (BDI-II) among postpartum women in Kedah, North West of Peninsular Malaysia. *Malays Journal of Medical Science*, *11(2),* 19-25.

Manson, S. M. (1995). Culture and major depression: current challenges in the diagnosis of mood disorders. *Psychiatric Clinics of North America*, *18,* 487-501.

Mantzicopoulos, P. Y., French, B. F., & Maller, S. J. (2004). Factor structure of the pictorial scale of perceived competence and social acceptance with two pre-elementary samples. *Child Development*, *75*, 1214–1228.

Maser, J. D., & Cloninger, C. R. (1990). *Comorbidity of mood and anxiety disorders*. Washington DC: American Psychiatric.

Marsella, A. J. (1980). Depressive experience and disorder across cultures. In H. Triandis & J. Draguns (Eds.), *Handbook of cross-cultural psychology: Vol. 6 Mental health* (pp. 237-289). Boston, MA: Allyn & Bacon.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cut-off values for fit indexes and dangers in overgeneralizing Hu and Bentler's findings. *Structural Equation Modeling, 11,* 320-341.

Marsh, H. W., Hey, J., & Roche, L. A. (1997). Structure of physical self-concept: Elite athletes and physical education students. *Journal of Educational Psychology*, *89,* 369–380.

Martens, E. J., Hoen, P. W., Mittelhaeuser, M., de Jonge, P., & Denollet, J. (2010). Symptom dimensions of post-myocardial infarction depression, disease severity and cardiac prognosis. *Psychological Medicine: A Journal of Research in Psychiatry & the Allied Sciences*, *40(5),* 807-814.

Martin, C. S., Pollock, N. K., Bukstein, O. G., & Lynch, K. G. (2000). Interrater reliability of the SCID alcohol and substance use disorders section among adolescents. *Drug & Alcohol Dependence*, *59*, 173–176.

Matsumoto, D. (1997). *Culture and modern life*. Pacific Grove, CA: Brooks/Cole.

McPherson, L. S., & Lakey, B. (1993). Content-overlap inflates the relation between dysphoria and negative cognition. *Journal of Personality Assessment*, *60*, 411-417.

Meiring, D., Van de Vijver, F. J. R., & Rothmann, S. (2006). Bias in an adapted version of the 15FQ+ in South Africa. *South African Journal of Psychology*, *36,* 340-356.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.

Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In A. Sayer & L. Collins (Eds.), *New methods for the analysis of change* (pp. 203–240). Washington, DC: American Psychological Association.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18,* 5–11.

Metalsky, G. I., Joiner, T. E., Jr., Hardin, T. S., & Abramson, L. Y. (1993). Depressive reactions to failure in a naturalistic setting: A test of the hopelessness and self-esteem theories of depression. *Journal of Abnormal Psychology, 102*, 101 - 109.

Mezzich, J. E., & Caracci, G. (2008). *Cultural formulation: A reader for psychiatric diagnosis.* New York, NY: Jason Aronson.

Miles, A., McManus, C., Feinmann, C., Glover, L., Harrison, S., & Pearce, S. (2001). The factor structure of the BDI in facial pain and other chronic pain patients: A comparison of two models using confirmatory factor analysis. *British Journal of Health Psychology*, *6(2),* 179-196.

Millsap, R. E. (2005). Four unresolved problems in studies of factorial invariance. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics* (pp. 153–172). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mollica, R. F., Wyshak, G., de Marneffe, D., Khuon, F., & Lavelle, J. (1987). Indochinese versions of the Hopkins Symptom Checklist-25: A screening instrument for the psychiatric care of refugees. *American Journal of Psychiatry*, *144(4),* 497–500.

Morley, T. E., & Moran, G. (2011). The origins of cognitive vulnerability in early childhood: Mechanisms linking early attachment to later depression. *Clinical Psychology Review*, *31*, 1071-1082.

Morley, S., Williams, A. C., & Black, S. (2002). A confirmatory factor analysis of the Beck Depression Inventory in chronic pain. *Pain, 99(1),* 289-298.

Moss, P. A. (1998). Recovering a dialectical view of rationality. *Social Indicators Research*, *45*, 55-67.

Mossakowski, K. (2008). Dissecting the influence of race, ethnicity, and socioeconomic status on mental health in young adulthood. *Research on Aging, 30,* 649-671.

Mukhtar, F., & Tian, P. O. (2008). Exploratory and confirmatory factor validation and psychometric properties of Beck Depression for Malays (BDI-Malay) in Malaysia. *Malaysian Journal of Psychiatry EJournal*.

Markus, H. R., & Kitayama, S. (1994). The cultural construction of self and emotion: Implications for social behavior. In S. Kitayama & H. R. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 89-130). Washington, DC: American Psychological Association.

Murray, C. J. L., & Lopez, A. D. (1996). *The global burden of disease: A comprehensive assessment of disability from diseases, injuries, and risk factors in 1990 and projected to 2020.* Cambridge, MA: Harvard University.

Nielsen, C., & Williams, T. A. (1980). Depression in ambulatory medical patients. *Archives of General Psychiatry*, *37*, 999-1004.

Nolen-Hoeksema, S. (1987). Sex differences in unipolar depression: Evidence and theory. *Psychology Bulletin*, *101*, 259–282.

Nolen-Hoeksema, S. (1990). *Sex differences in depression*. Stanford, CA: Stanford University.

Nolen-Hoeksema, S., Larson, J., & Grayson, C. (1999). Explaining the gender difference in depressive symptoms. *Journal of Personality & Social Psychology*, *77*, 1061–1072.

Nolen-Hoeksema, S., & Girgus, J. S. (1994). The emergence of gender differences in depression during adolescence. *Psychological Bulletin*, *115*, 424-443.

Oliver, J. M., & Simmons, M. E. (1985). Affective disorders and depression as measured by the Diagnostic Interview Schedule and the Beck Depression Inventory in an unselected adult population. *Journal of Clinical Psychology*, *41,* 469-477.

Orth, U., Robins, R. W., & Meier, L. L. (2009). Disentangling the effects of low self-esteem and stressful events on depression: Findings from three longitudinal studies. *Journal of Personality & Social Psychology*, *97,* 307-321.

Orth, U., Robins, R. W., & Roberts, B. W. (2008). Low self-esteem prospectively predicts depression in adolescence and young adulthood. *Journal of Personality & Social Psychology*, *95*, 695-708.

Orth, U., Robins, R. W., Trzesniewski, K. H., Maes, J., & Schmitt, M. (2009). Low self-esteem is a risk factor for depressive symptoms from young adulthood to old age. *Journal of Abnormal Psychology*, *118*, 472-478.

Osman, A., Barrios, F., Gutierrez, P. M., Williams, J. E., & Bailey, J. (2008). Psychometric properties of the Beck Depression Inventory–II in nonclinical adolescent samples. *Journal of Clinical Psychology*, *64*, 83-102.

Osman, A., Downs, W. R., Barrios, F. X., Kopper, B. A., Gutierrez, P. M., & Chiros, C. E. (1997). Factor structure and psychometric characteristics of the Beck Depression Inventory II. *Journal of Psychopathology & 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*, 120-132.

Owen, K. (1992). *Test-item bias: Methods, findings and recommendations*. Pretoria, South Africa: Human Sciences Research Council, Group: Education.

Peacock, R. S., & Theron, A. (2007). Identity development of the Incarcerated adolescent with particular reference to prison gang membership. *Acta Criminologica, 20(3),* 61-74.

Peltzer, K., & Habil, M. (1998). Psychology and health in South Africa. *Inaugural lecture for professorship of the Department of Psychology,* University of the North: Polokwane, South Africa.

Penley, J. A., Wiebe, J. S., & Nwosu, A. (2003). Psychometric properties of the Spanish Beck Depression Inventory-II in a medical sample. *Psychological Assessment, 15*, 569-577.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oaks, CA: Sage.

Pillay, A., & Sargent, C. (1999). Relationship of age and education with anxiety, depression, and hopelessness in a South African community sample. *Perceptual & Motor Skills*, *89,* 881-884.

Ployhart, R. P., & Oswald, F. L. (2004). Applications of mean and covariance structure analysis: Integrating correlational and experimental approaches. *Organizational Research Methods*, *7*, 27-65.

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

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, *87*, 517-529.

Rao, U., & Chen, L. (2009). Characteristics, correlates, and outcomes of childhood and adolescent depressive disorders. *Dialogues in Clinical Neuroscience*, *11,* 45-62.

Razali, S. M., & Hasanah, C. I. (1999). Somatization of depression: Prevalence and characteristics features. *International Medical Journal*, *6,* 27-31.

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.

Roberts, R. E., Lewinsohn, P. M., & Seeley, J. R. (1991). Screening for adolescent depression: A comparison of depression scales. *Journal of the American Academy of Child & Adolescent Psychiatry*, *30*, 58-66.

Roberts, J. E., & Monroe, S. M. (1992). Vulnerable self-esteem and depressive symptoms: Prospective findings comparing three alternative conceptualizations. *Journal of Personality & Social Psychology, 62*, 804-812.

Rodriguez-Gomez, J. R., Davila-Martinez, M. G., Collazo-Rodriguez, L. C. (2006). Factor structure of the Beck Depression Inventory-Second Edition (BDI-II) with Puerto Rican elderly. *Puerto Rico Health Sciences Journal*, *25,* 127-132.

Rodríguez, J.R., Joglar, P., & Dávila, M.G. (2005). The relationship between depressive symptomatology and high blood pressure in Hispanic elderly. *American Journal of Health Studies*, *20*, 3.

Rosado, M. (1995). *Depresión en ancianos con y sin apoplejía(s) cerebral(es) o amputación(es) que reciben tratamiento de rehabilitación física: Un estudio piloto en Puerto Rico* [Disertación]. San Juan (PR): Centro Caribeño de Estudios Postgraduados. Abstract retrieved from http://prhsj.rcm.upr.edu/index.php/prhsj/article/viewFile/315/206

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University.

Rowland, S. M., Lam, C. S., & Leahy, B. (2005). Use of the Beck Depression Inventory-II (BDI-II) with persons with traumatic brain injury: Analysis of factorial structure. *Brain Injury, 19(2),* 77–83.

Rushton, J.P., Skuy, M., & Fridjhon, P. (2003). Performance on Raven's Advanced Progressive Matrices by African, Indian, and White engineering students in South Africa. *Intelligence*, *31*, 123–137.

Santor, D. A., Ramsay, J. O., & Zuroff, D. C. (1994). Nonparametric item analyses of the Beck Depression Inventory: Evaluating gender item bias and response option weights. *Psychological Assessment*, *6*, 255-270.

Satorra, A., & Bentler, P. M. (1988). *Scaling corrections for chi-square statistics in covariance structure analysis.* ASA 1988 Proceedings of the Business and Economic Statistics, Section (308-313). Alexandria, VA: American Statistical Association. Retrieved from http://www.amstat.org/meetings/csp.cfm

Schuch, J. J., Roest, A. M., Nolen, W. A., Penninx, B. W., & de Jonge, P. (2014). Gender differences in major depressive disorder: results from the Netherlands study of depression and anxiety. *Journal of Affective Disorders*, *156*,156-163

Schwab, J. J., Bialow, M., Brown, H. M., & Holzer, C. E. (1967). Diagnosing depression in medical inpatients*. Annals of Internal Medicine*, *67,* 695–707.

Segal, D. L., Coolidge, F. L., Cahill, B. S., & O'Riley, A. A. (2008). Psychometric properties of the Beck Depression Inventory II (BDI-II) among community-dwelling older adults. *Behavior Modification, 32(1),* 3–20.

Seignourel, P. J., Green, C., & Schmitz, J. M. (2008). Factor structure and diagnostic efficiency of the BDI-II in treatment-seeking substance users. *Drug & Alcohol Dependence*, *93*, 271–278.

Shafer, A. B. (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton and Zung. *Journal of Clinical Psychology*, *62*, 123-146.

Shahani. C., Dipboye, R. L., & Phillips, A. (1990). Global self-esteem as a correlate of work-related attitudes: A question of dimensionality. *Journal of Personality Assessment*, *54(1 & 2),* 276-288.

Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The Self-Efficacy Scale: Construction and validation. *Psychological Reports*, *51,* 663-671.

Siegel, J. M., Yancey, A. K., Aneshensel, C. S., & Schuler, R. (1999). Body image, perceived pubertal timing, and adolescent mental health. *Journal of Adolescent Health, 25,* 155–165.

Silverstein, B. (1999). Gender difference in the prevalence of clinical depression: The role played by depression associated with somatic symptoms. *American Journal of Psychiatry*, *156*, 480-482.

Sivo, S. A., Fan, X. T., Witta, E .L., & Willse, J. T. (2006). The search for 'optimal' cutoff properties: Fit index criteria in structural equation modeling. *The Journal of Experimental Education*, *74(3),* 267-289.

Sörbom, D. (1978). An alternative to the methodology for analysis of covariance. *Psychometrika*, *43*, 381–396.

Statistics South Africa. (2001). Census 2001 by province, gender, disability and population group. Retrieved from http://www.statssa.gov.za/extract.htm

Steele, G. I. (1996). *The feasibility of translating the Beck Depression Inventory and the Beck Anxiety Inventory into Xhosa.* Unpublished Honours Dissertation. Rhodes University: Grahamstown.

Steele, G. I., & Edwards, D. J. A. (2008). Development and validation of the Xhosa translations of the Beck Inventories: 2. Item analyses and internal consistency. *Journal of Psychology in Africa*, *18(2),* 217-226.

Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1999). Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *Journal of Clinical Psychology*, *55(1),* 117-128.

Steer, R. A., Beck, A. T., Riskind, J. H., & Brown, G. (1986). Differentiation of depressive disorders from generalized anxiety by the Beck Depression Inventory. *Journal of Clinical Psychology*, *42(3),* 475–478.

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

Steer, R. A., Kumar, G., Ranieri, W., & Beck, A. T. (1998). Use of the Beck Depression Inventory—II with adolescent psychiatric outpatients. *Journal of Psychopathology & Behavioral Assessment*, *20*, 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(3),* 311-318.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180.

Stewart, S. M., Lam, T. H., Betson, C., & Chung, S. F. (1999). Suicide ideation and its relationship to depressed mood in a community sample of adolescents in Hong Kong. *Suicide & Life-Threatening Behavior*, *29*, 227-240.

Storch, E. A, Roberti, J. W, & Roth, D. A. (2004). Factor structure, concurrent validity, and internal consistency of the Beck Depression Inventory–Second edition in a sample of college students. *Depression & Anxiety, 19*, 187-189.

Storm, K., & Rothmann, S. (2003). A psychometric analysis of the Maslach Burnout Inventory-General Survey in the South African police service. *South African Journal of Psychology*, *33*, 219-226.

Swanepoel, B. A., & Krüger, C. (2011). Revisiting validity in cross-cultural psychometric test development: A systems-informed shift towards qualitative research designs. *South African Journal of Psychiatry*, *17(1),* 10-15.

Syed, H. R., Zachrisson, H. D., Dalgard, O. S., Dalen, I., & Ahlberg, N. (2008). Concordance between Hopkins Symptom Checklist (HSCL-10) and Pakistan Anxiety and Depression Questionnaire (PADQ), in a rural self-motivated population in Pakistan. *BMC Psychiatry*, *8(59),* 1-12.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Upper Saddle River, NJ: Pearson, Allyn & Bacon.

Teja, J. S., Narang, R. L., & Aggarwal, A. K. (1971). Depression across cultures. *British Journal of Psychiatry*, *119*, 253-260.

Tems, C. L., Stewart, S. M., Skinner, J. R., Hughes, C. W., & Emslie, G. (1993). Cognitive distortions in depressed children and adolescents: Are they state dependent or trait-like? *Journal of Clinical Child Psychology*, *22,* 316–326.

Thombs, B. D., Ziegelstein, R. C., Beck, C. A., & Pilote, L. (2008). A general factor model for the Beck Depression Inventory-II: Validation in a sample of patients hospitalized with acute myocardial infarction. *Journal of Psychosomatic Research*, *65(2),* 115-121.

Tomlinson, M., Grimsrud, A.T., Stein, D. J., Williams, D. R., & Myer, L. (2009). The epidemiology of major depression in South Africa: Results from the South African stress and health study. *South African Medical Journal*, *99(5),* 367-373.

Tyler, G. (2002). *A review of the 15FQ+ Personality Questionnaire*. Pulloxhill, UK: Psychometrics.

Walker, R. L., & Bishop, S. (2005). Examining a model of the relation between religiosity and suicidal ideation in a sample of African American and white college students. *Suicide & Life-Threatening Behavior*, *35*, 630-639.

Wang, Y. P., Andrade, L. H., & Gorenstein, C. (2005). Validation of the Beck Depression Inventory for a Portuguese-speaking Chinese community in Brazil. *Brazilian Journal of Medical & Biological Research*, *38(3),* 399-408.

Ward, L. C. (2006). Comparison of factor structure models for the Beck Depression Inventory-II. *Psychological Assessment*, *18(1),* 81-88.

Westaway, M. S., & Wolmarans, L. (1992). Depression and self-esteem: Rapid screening for Depression in Black, low literacy, hospitalized Tuberculosis patients. *Social Science & Medicine*, *35*, 1311-1315.

Whisman, M. A., Judd, C. M., Whiteford, N. T., & Gelhorn, H. L. (2013). Measurement invariance of the Beck Depression Inventory-Second Edition (BDI-II) across gender, race, and ethnicity in college students. *Assessment*, *20(4),* 419–428.

Whisman, M. A., & Kwon, P. (1993). Life stress and dysphoria: The role of self-esteem and hopelessness. *Journal of Personality & Social Psychology*, *65*, 1054–1060.

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.

Wiebe, J. S., & Penley, J. A. (2005). A psychometric comparison of the Beck Depression Inventory-II in English and Spanish. *Psychological Assessment*, *17,* 481-485.

World Health Organisation. (2001). *World health report 2001. Mental health resources in the world.* Geneva: World Health Organisation.

Wu, P.-C. (2009). Differential functioning of the Chinese version of Beck Depression Inventory-II in adolescent gender groups: Use of a multiple-group mean and covariance structure model. *Social Indicators Research*. **DOI**: 10.1007/s11205-009-949-0

Wu, P.-C. (2010). Measurement invariance and latent mean differences of the Beck Depression Inventory II across gender groups. *Journal of Psychoeducational Assessment*, *28*, 551-563.

Wu, P.-C., & Chang, L. (2008). Psychometric properties of the Chinese version of Beck Depression Inventory II (BDI-II-C) using Rasch model. *Measurement & Evaluation in Counseling & Development*, *41,* 13-31.

Wu, P.-C., & Huang, T.-W. (2012). Gender-related invariance of the Beck Depression Inventory II for Taiwanese adolescent samples. *Assessment*. **DOI**: 10.1177/1073191112441243

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods, 3*, 4-69.

van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, *1*, 89–99.

Van Eeden, R., & Mantsha, T. R. (2007). Theoretical and methodological considerations in the translation of the 16-PF-5 into an African language. *South African Journal of Psychology*, *37,* 62-81.

Vanheule, S., Desmet, M., Groenvynck, H., Rosseel, Y., & Fontaine, J. (2008). The factor structure of the Beck Depression Inventory-II: An evaluation. *Assessment*, *15*, 177-187.

Van Wyk, R., Boshoff, A. B., & Owen, J. H. (1999). Construct validity of psychometric instruments developed in the United States, when applied to professional people in South Africa. *South African Journal of Economic & Management Sciences*, *1*, 1-72.

Viljoen, J. L., Grant, L. I., Griffiths, S., & Woodward, T. S. (2003). Factor structure of the Beck Depression Inventory–II in a medical outpatient sample. *Journal of Clinical Psychology in Medical Settings*, *10*, 289–291.

Vredenburg, K., Krames, L., & Flett, G. L. (1986). Sex differences in the clinical expression of depression. *Sex Roles*, *14,* 37–49.

Yuan, K.-H., & Bentler, P. M. (2000). Robust mean and covariance structure analysis through iteratively reweighted least squares. *Psychometrika, 65*, 43–58.

Zanarini, M. C., & Frankenburg, F. R. (2001). Attainment and maintenance of reliability of Axis I and II disorders over the course of a longitudinal study. *Comprehensive Psychiatry*, *28*, 467–480.

Zanarini, M. C., Skodol, A. E., Bender, D., Dolan, R., Sanislow, C., Schaefer, E., et al. (2000). The collaborative longitudinal personality disorder study: Reliability of axis I and II diagnoses. *Journal of Personality Disorders*, *14*, 291–299.

Zumbo, B. D. (2009). Validity as contextualized and pragmatic explanation, and its implications for validation practice. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 65–82). Charlotte, NC: IAP.

APPENDICES

Appendix A: Demographic details

[Please note: The researcher does not have the rights to publish the scales/psychological measures used in this study. Therefore, the scales are not appended.]

Section 1: Demographic questionnaire

1. How old are you? ………….yrs. old.
2. Which "race" or ethnic group do you belong to? (Choose one answer)

Black		Coloured		Asian		White	

   N.B. Please note that this item is used for research purposes only.

3. Are you male or female? (Choose one answer)

Male		Female	

4. What is your family's estimated gross income per year (please tick the appropriate box)?

Less than R20 000	
R21 000 - R40 000	
R41 000 - R60 000	
R61 000 - R80 000	
R81 000 - R100 000	
R101 000 or more	

Appendix B: **Information for participants, consent form and ethical clearance from the Universities of Limpopo and Pretoria**

PART I:

UNIVERSITY OF LIMPOPO
Turfloop Campus
FACULTY OF HUMANITIES
SCHOOL OF SOCIAL SCIENCES
Department of Psychology

INFORMATION FOR PARTICIPANTS

PROJECT TITLE: VALIDATION OF THE BDI-II IN SOUTH AFRICA

PROJECT LEADER: MAKHUBELA MALOSE SILAS

1. You are invited to participate in the following research project:

   "VALIDATION OF THE BDI-II IN SOUTH AFRICA"

2. Participation in the project is completely voluntary and you are free to withdraw from the project (without providing any reasons or consequences) at any time.

3. It is possible that you might not personally experience any advantages during the project, although the knowledge that may be accumulated through the project might prove advantageous to others.

4. You are encouraged to ask any questions that you might have in connection with this project at any stage. The project leader and her/his staff will gladly answer your question. They will also discuss the project in detail with you.

5. There are no known consequences of completing a questionnaire about Depressive symptomatology. However, individuals who have experienced clinical forms of depression may react apprehensively; being sensitive to completing questions about situations/symptoms that were not particularly comfortable for them.

6. Should you at any stage feel unhappy, uncomfortable or is concerned about the research, please **contact Ms Noko Shai-Ragoboya at the University of Limpopo, Private Bag X1106, Sovenga, 0727, tel: 015 268 2401**.

PART II:

UNIVERSITY OF LIMPOPO
Turfloop Campus
FACULTY OF HUMANITIES
SCHOOL OF SOCIAL SCIENCES
Department of Psychology

CONSENT FORM

PROJECT TITLE: VALIDATION OF THE BDI-II IN SOUTH AFRICA

PROJECT LEADER: MAKHUBELA M.S.

I, ____ hereby voluntarily consent to participate in the following project:

"VALIDATION OF THE BDI-II IN SOUTH AFRICA"

I realise that:

1. The study deals with the evaluation of the degree to which items or subtests of the Beck Depression Inventory-II (BDI-II) have equal meaning across groups of examinees in South Africa.

2. The procedure envisaged may hold some risk for me that cannot be foreseen at this stage.

3. The Ethics Committee of the University of Limpopo has approved that individuals may be approached to participate in the study.

4. The research project, i.e. the extent, aims and methods of the research, has been explained to me.

5. The project sets out the risks that can be reasonably expected as well as possible discomfort for persons participating in the research, an explanation of the anticipated advantages for myself or others that are reasonably expected from the research and alternative procedures that may be to my advantage.

6. I will be informed of any new information that may become available during the research that may influence my willingness to continue my participation.
7. Access to the records that pertain to my participation in the study will be restricted to persons directly involved in the research.

8. Any questions that I may have regarding the research, or related matters, will be answered by the researcher/s.

9. If I have any questions about, or problems regarding the study, or experience any undesirable effects, I may contact a member of the research team or **Ms Noko Shai-Ragoboya.**

10. Participation in this research is voluntary and I can withdraw my participation at any stage.

11. If any medical problem is identified at any stage during the research, or when I am vetted for participation, such condition will be discussed with me in confidence by a qualified person and/or I will be referred to my doctor.

12. I indemnify the University of Limpopo and all persons involved with the above project from any liability that may arise from my participation in the above project or that may be related to it, for whatever reasons, including negligence on the part of the mentioned persons.

____________________________	____________________________
SIGNATURE OF RESEARCHED PERSON	SIGNATURE OF WITNESS
____________________________	____________________________
SIGNATURE OF PERSON THAT INFORMED THE RESEARCHED PERSON	SIGNATURE OF PARENT/GUARDIAN

Signed at_________________________ this ____ day of ____________ 20__

PART III: Ethical clearance and permission to access students

University of Limpopo
Research Development and Administration Department
Private Bag X1106, Sovenga, 0727, South Africa
Tel: (015) 268 2212, Fax: (015) 268 2306, Email:noko.monene@ul.ac.za

TURFLOOP RESEARCH ETHICS COMMITTEECLEARANCE CERTIFICATE

MEETING: **04 September 2013**

PROJECT NUMBER: **TREC/FHM/47/2013: PG**

PROJECT:

Title:	Validation of the BDI-II in South Africa
Researcher:	Mr MS Makhubela
Supervisor:	Dr S Mashegoane
Co-Supervisor:	N/A
Department:	Psychology
School:	Social Sciences
Degree:	PhD in Psychology

PROF TAB MASHEGO
CHAIRPERSON: TURFLOOP RESEARCH ETHICS COMMITTEE

The Turfloop Research Ethics Committee (TREC) is registered with the National Health Research Ethics Council, Registration Number: **REC-0310111-031.**

Note:

i) Should any departure be contemplated from the research procedure as approved, the researcher(s) must re-submit the protocol to the committee.

ii) The budget for the research will be considered separately from the protocol.

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES.

Finding solutions for Africa

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Humanities
Research Ethics Committee

25 June 2013

Dear Mr Makhubela

Project:	**Validation of BDI-II in South Africa**
Researcher:	**SM Makhubela**
Department:	**Psychology**
Reference:	**Staff research**

I have pleasure in informing you that the Registrar has formally given **approval** for the above study to be conducted at the University of Pretoria. Data collection may therefore commence.

Please note that this approval is based on the assumption that the research will be carried out along the lines laid out in the proposal. Should your actual research depart significantly from the proposed research, it will be necessary to apply for a new research approval and ethical clearance.

The Committee requests you to convey this approval to the researcher.

We wish you success with the project.

Sincerely

Prof. Sakhela Buhlungu
Chair: Research Ethics Committee
Faculty of Humanities
UNIVERSITY OF PRETORIA
e-mail: sakhela.buhlungu@up.ac.za

Research Ethics Committee Members: Prof S Buhlungu (Chair); Dr L Blokland; Prof M-H Coetzee; Dr JEH Grobler; Prof KL Harris; Ms H Klopper; Dr Sepetla Molapo; Dr G Pelaker; Dr C Panebianco-Warrens; Prof GM Spies; Prof E Taljard ; Dr FG Wolmarans, Dr P Wood