Graduate Theses and Dissertations

Graduate School

2011

Translation and Adaptation of the Center for Epidemiologic Studies-Depression (CES-D) Scale Into Tigrigna Language for Tigrigna Speaking Eritrean Immigrants in the United States

Mulubrhan Fisseha Mogos
University of South Florida, muler1978@yahoo.com

Follow this and additional works at: http://scholarcommons.usf.edu/etd

Part of the American Studies Commons, Nursing Commons, Psychiatric and Mental Health Commons, and the Social and Behavioral Sciences Commons

Translation and Adaptation of the Center for Epidemiologic Studies-

Depression (CES-D) Scale into Tigrigna Language for Tigrigna Speaking

Eritrean Immigrants in the United States

by

Mulubrhan F. Mogos

A dissertation submitted in partial fulfillment
of the requirements for the degree
Doctor of Philosophy
College of Nursing
University of South Florida

Co-Major Professor: Jason Beckstead, Ph.D.
Co-Major Professor: Mary E. Evans, Ph.D., RN, FAAN
Kevin E. Kip, Ph.D., FAHA
Roger A. Boothroyd, Ph.D.

Date of Approval:
May 31, 2011

Dedication

I dedicate this work to my family: my father, mother, sister, and brothers for their love,

endless support and encouragement.

Acknowledgments

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AGFI	Adjusted Goodness-of-Fit Statistics
CES-D	Center for Epidemiologic Studies Depression
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
DA	Depressive Affect
DALYs	Disability Adjusted Life Years
EPLF	Eritrean People Liberation Front
GFI	Goodness-of-Fit Statistics
HEART SCORE	HEART Strategies Concentrating on Risk Evaluation
IP	Interpersonal
ITC	Council of International Test Commissions
IRB	Institutional Review Board
LISRE	Linear Structural Relations
MGCFA	Multi-Group Confirmatory Factor Analysis
MIs	Modification Indices
ML	Maximum Likelihood
MMPI	Minnesota Multiphase Personality Inventory
NFI	Normed Fit Index
NIMH	National Institute of Mental Health
PA	Positive Affect
PTSD	Post Traumatic Stress Disorder
RMR	Root mean Square Residual
RMSEA	Root mean Square Error of Approximation
SD	Standard Deviation
SEM	Structural Equation Modeling
SRMR	Standardized Root mean Square Residual
SV	Somatic Vegetative
SES	Socio Economic Status
UNHCR	United Nations Higher Commission for Refugees
US	United States
USF	University of South Florida
USCR	United States Commission for Refugees
USPSTF	United States Preventive Services Task Force
WHO	World Health Organization
YLDs	Years Lived with Disability

Abstract

Depression is one of the oldest known mental health conditions. It is acknowledged to be a global health problem that affects people from any culture or ethnic group. The prevalence of depression widely varied across countries and cultures. The cross-cultural relevance of the concept of depression, its screening or diagnosis, and cultural equivalence of items used to measure symptoms of depression has been area of research interest. Differences in prevalence rates in depression have been suggested as being due to research artifacts, such as use of instrument developed for one culture to another culture. With the current trend of globalization and increased rate of immigration, the need for measurement scales that can be used cross-culturally is becoming essential. Translation and adaptation of existing tools to different languages is time saving and cost effective than developing a new scale. The Center for Epidemiologic Studies Depression Scale [CES-D; (Radloff, 1977)] has been widely used as a screening tool for depressive symptoms in community and clinical settings. It has been widely accepted and translated to multiple languages and its measurement equivalence tested across groups. This study was designed as a mixed method study. The purpose of this study was three fold: (a) translate and adapt the CES-D scale into Tigrigna Language for use by Tigrigna speaking Eritrean immigrants in the United States using the forward backward translation and cognitive interview techniques (b) test the psychometric properties of the Tigrigna version CES-D scale using confirmatory factor analysis under the framework of

structural equation modeling and (c) test measurement equivalence of the scale by comparing data collected from 253 Eritrean immigrants using the Tigrigna version CES-D scale with a secondary data collected from 1918 non Eritrean US citizens using the English version CES-D scale in a separate study. The baseline four factor CES-D scale model originally suggested for the general population fitted the data from both samples. The fit indices for the Tigrigna sample were ($\chi^2 = 299.87$, df = 164, RMSEA = .06, SRMR = .06, GFI = .89, and CFI = .98) and for the English sample ($\chi^2 = 1496.81$, df = 164, RMSEA = .07, SRMR = .04, GFI = .92, and CFI = .98). The Multi-group confirmatory factor analysis showed reasonably adequate fit ($\chi^2_{(328)} = 1796.68$, RMSEA= .07, SRMR = .06, GFI = .89, CFI = .98). Fourteen of the 20 CES-D items were invariant across the two samples suggesting partial metric invariance. Partial full factor invariance was also supported. In conclusion, the findings of this study provide adequate evidence in support of the applicability of the four factor CES-D scale for measuring depressive symptoms in Tigrigna speaking Eritrean immigrants/refugees in the United States.

Chapter One

Introduction

The increasingly diverse nature of the population in the United States is drawing researchers' interest to minority groups such as immigrants. Doing research with minority groups requires valid and reliable measurement tools. The interest and awareness of cross-cultural aspects of illness and wellness has been increasing over time. With such an increase in awareness of cross-cultural aspects of illness and wellness, the consideration of linguistic translations and adaptation of culturally appropriate research tools for use in languages other than the source language is also growing significantly (Anderson, Aaronson, & Wilkin, 1993; Wiesinger et al., 1999). Through years of hard work by various researchers, many instruments have been developed and validated to measure physical symptoms, functional status, psychological state, and social interactions. The majority of these measurement tools are developed in English speaking countries (Guillemin, Bombardier, & Beaton, 1993). But even in these countries the use of those tools might not be appropriate for immigrant/refugee populations sometimes forcing researchers to exclude immigrant/refugee population from their studies. The consequence of such exclusion could cause systematic bias which makes generalization of findings impossible.

The use of existing tools which are already validated is time saving and convenient. Such use will save the amount of money and time to develop a new tool and most importantly it helps to gather more information on the psychometric properties of existing tools in various populations. Use of a tool developed for one culture in a different culture requires examining its cultural relevance for the culture it is being planned to employ. This process requires adaptation of the tool from the source language into the target language (i.e., the language into which the tool is to be translated) without losing its original meaning in the process. Instrument adaptation encompasses conceptual meanings in the source language within the context of the translation process. This will help to maintain construct equivalence and content representation across the two languages. Therefore, cross cultural adaptation of health status measurement tools for a new culture and language requires a systematic methodology to achieve equivalence between the original or source tool and target version of the tool. The objective of this study is to translate and adapt the 20- item Center for Epidemiologic Studies-Depression (CES-D) scale into Tigrigna language and test its psychometric properties in Tigrigna speaking Eritrean immigrants/refugees in the United States.

The terms "refugees" and "immigrants" have been used interchangeably in the literature. Although a refugee movement is similar to an immigrants' movement, the former is used for individuals who left their territory because of political events rather than economic interest. In this study the term immigrants is used to describe all Eritreans who left Eritrea during the long war for independence with Ethiopia and those who continue to flee from militarism, political repression, fear of another round of boarder

conflict with Ethiopia, and those who came through programs like diversity visa, family reunion, and education and currently reside in the United States.

Because of its post colonial era annexation with its neighbor Ethiopia in 1962, Eritrea as a country is less familiar to the rest of the world. Therefore, it often requires some detailed background information to introduce readers to this small new east African nation in the horn of Africa, along the coast of the Red Sea. Like the formation of almost all nations in Africa, Eritrea is a product of modern colonialism by Italians from 1885 to 1941. The unique thing about the history of colonialism in Eritrea is that it is the only African country denied of its right for independence and forcefully annexed by Ethiopia making Eritrea one of its provinces. The forceful annexation of Eritrea in 1962 sparked one of the longest wars for independence in the history of Africa and Eritrea gained independence in 1991. Between 1991 and 1998, war, destruction, and oppression seemed to be replaced with a bright future with promises of development and prosperity specifically for Eritrea and for the horn of Africa in general. However; this hope didn't last long and in 1998 another border conflict erupted between Eritrea and Ethiopia, creating new problems for the people of Eritrea in particular and the horn of Africa in general.

The 30 year long struggle for independence, the recent boarder conflict with neighboring Ethiopia, and the existing political & socioeconomic situation forced and is still forcing Eritrean men and women to leave their villages and cities and seek shelter in neighboring countries, mainly the Sudan. The first recognizable influx of Eritrean refugees to the Sudan began in March 1967 (Gaim, 1987). This initial entry of Eritrean

refugees to neighboring country Sudan continued with other mass influxes between late sixties and late seventies.

To my knowledge there is no reliable documented statistics on the actual number of Eritrean immigrants worldwide. However, different sources have estimated the number to be about one million which is about one-quarter of the Eritrean population (Hepner & Conrad, 2005). The first Eritrean immigrant students and workers came to the US following the Immigration Act of 1965. However, a resettlement program of Eritreans from refugee camps in Sudan following the Refugee Act of 1980 was the time Eritreans began to resettle in the United States in relatively large numbers (Gaim, 1985). According to a report by the United States Commission for Refugees, from 1988 to 1992 alone 21, 901 Eritreans and Ethiopians entered the United States (USCR, 2001). Woldemichael (1998) estimated the number of individuals resettled between 1975 and 1994 to be about 33,200. This statistics doesn't differentiate between Eritreans and Ethiopians. Since the mid 1990s the number of Eritrean immigrants in the US has also been growing fast because of newly arriving Eritreans through diversity program, family reunion, and recently also asylum seeking immigrants. According to reports from Eritrean Embassy in the United States, in 1993 about 20, 000 Eritrean immigrants (18 years old and above) participated in the national referendum for independence. According to some sporadic reports from Eritrean websites, a minimum of 150, 000 Eritreans are believed to reside in the United States at present. Many of these immigrants have firsthand experience of violence while being in Eritrea or in refugee camps.

Statement of the Problem

In the year 2002, the United States Preventive Services Task Force (USPSTF) recommended screening of depression in primary care settings (Pignone et al., 2002).The cross-cultural differences in the incidence and prevalence of depressive symptoms are not necessarily genuine variations. Those differences are far from true biological differences and could be reflections of multiple causes including perception of interpersonal interactions (Boutin-Foster, 2008). For example Cole et al. (2000), found out that African-Americans who scored high on the CES-D scale were more likely to report that other people were unfriendly toward them or that they felt disliked by others. Barnes et al. (2004) also reported strong correlation between perception of maltreatment and rejection with depressive symptom. The growing number of immigrants in the United States has promoted an interest in validating various screening instruments among different groups of immigrants. Immigrant populations in the United States are extremely diverse in terms of culture, language, socioeconomic status (SES), pre immigration experience, and even post immigration experiences. For example immigrants from Eritrea do have different experience than those from other African countries such as Nigeria. Despite such variation, no study has been done to investigate the mental health status of Eritrean immigrants who are among the first group of immigrants to the United States, and arguably among immigrants who experienced multiple traumatic events prior to and during their immigration process.

Eritrean immigrants are among the first group of immigrants from Africa following the immigration act of 1965 in the United States. In the past two to three decades they were able to establish strong community based organizations through

5

gatherings in church, national holidays, and national youth sport festivals. These

community gatherings have helped them to establish social support system. However,

through my personal experience and informal meetings with Eritrean immigrants in the

US, I was able to notice mental health related issues as witnessed by the presence of

immigrants with severe mental disability, sporadic incidences of suicide, and homicide.

Unfortunately, despite being at risk of mental health problems because of their

experiences of violence in Eritrea, refugee camps, and through their journey all the way

to the US, nothing has been done to study mental health problems in Eritrean Immigrants.

This study is the first step to translate and adapt the Center for Epidemiologic Studies-

Depression (CESD [Radloff, 1977]) scale into Tigrigna language to study the prevalence

of depressive symptoms in Eritrean immigrants in the United States.

Significance of the Study

Cross-cultural research is a preferred strategy in studying health problems like

depression because conceptions related to depression are often ethnocentric (Marsella,

1981; Kleinman, 1982). Hence, cross-cultural approach helps to investigate universal and

specific causes, manifestations, and experiences of depressive symptoms. Traumatic

experiences in refugees are often the results of events that continue to happen from the

native country and into the host country. Examples of traumatic experiences often faced

by immigrants include: political repression, detention, torture, terror, battlefield

experiences, disappearance of relatives and friends, separation and loss of families and

friends, hardships during the flight or in refugees camps, and loss of social status. Often

immigrants are a more vulnerable group compared to members of their respective

population at their country of origin and hence are at greater risk of developing mental health problems like depression (Gaim, 2009).

Each culture has its own way of defining and manifesting health problems making the detection and management of such conditions more difficult in immigrants because often the cultural background of health professionals in the host nation do not share the immigrants' cultural background. Most immigrants from Eritrea speak little or no English. Hence difficulty in communication can cause uncertainty about the meaning of both written and verbal interactions.

A Patient centered health care system with emphasis on early detection and prevention of health problems is becoming the focus of global health care system. This kind of health care system allows individuals to be involved actively by accurately reporting their feelings, functional abilities, symptoms, and perceptions. Therefore the type of measurement tool used to measure an individual's health condition should deliver the assessment items to the individual patient without ambiguity to gather relevant information about the individual's health condition. Most health status measurement tools at present have been developed in English and their psychometric property also tested in groups for whom the tool was intended making cross-cultural use of such tools problematic. A Tigrigna version of the CESD- scale in addition to its use for Eritrean immigrants, potentially it could be used also by Tigrigna speaking Ethiopian immigrants who basically share a very similar culture.

Several issues underscore the importance of translating and adapting the CES-D scale into Tigrigna language for use in Eritrean immigrants in the United States: studies show that language used to administer a measurement tool affect the way people answer

questions, there is considerable variation between different immigrant groups and a study done in one immigrant group about depression cannot be generalized to another group of immigrants, and in Eritrean culture, generally people are reluctant to talk about mental health issues like depression as there is social stigma associated to it. Therefore, mental health issues like depression remain obscured until they become serious. Hence, translation and adaptation of screening tool would help to identify immigrants with symptoms of depression before the symptoms get worse (Mills & Henretta, 2001). Developing Tigrigna version of CES-D scale will help to assess the extent of depressive symptoms in Eritrean Immigrants in the United States which will further provide basis for facilitating comparison of depressive symptoms with other immigrants and endogenous citizens of the United States.

The translation and adaptation of the CES-D scale into Tigrigna language for use in the Eritrean immigrant community requires careful design and through validation to ensure that cognitive concepts in the original tool can be appropriately translated and applied to the Eritrean culture. This study will be guided by set of rigorous guidelines (see method section) to develop Tigrigna version of CES-D scale which has sound psychometric properties to measure depression symptoms in Tigrigna speaking Eritrean immigrants and potentially also Tigrigna speaking Ethiopian immigrants in the United States.

The selection of the CES-D scale among other available depression symptom measurement tools is based on the satisfactory psychometric properties of the tool in cross cultural studies. The tool has been translated to number of languages including Chinese, Greek, Korean, Armenian, Arabic, and Spanish and proven to have good

psychometric properties. The twenty items used in this tool are short and relatively easier to understand. Moreover, most items were rated well in terms of their comprehension and cultural equivalence by the principal investigator and selected individuals from Eritrean immigrants.

Research Questions

In this study the following three questions will be answered:

- Research question one: Do items in the Tigrigna version demonstrate comparable loadings with items in the English version?

- Research question two: Does the Tigrigna version CES-D scale replicate the psychometric properties of the original CES-D scale.

- Research question three: Does the Tigrigna version CES-D scale replicate the four factor model initially suggested by Radloff (1977)?

Specific Aims

The specific aims of this study will be:

1. To translate and adapt the CES-D scale from original (English) language to the target (Tigrigna) language for use in Eritrean immigrants in the Unites States

2. To test the psychometric properties of the new Tigrigna version CES-D scale

Summary

Eritrea is a small new country in the horn of Africa along the cost of the Red Sea. The population of Eritrea is estimated to be around four million and about one-quarter of the total population is believed to be in diaspora during the 30 years long war for independence. In recent years following the eruption of new boarder conflicts with Ethiopia in 1998, thousands of Eritreans fled the country to escape political repression,

military conscription, and economic hardship increasing the number of Eritreans living abroad even more. The United States is one of the home countries for Eritrean immigrants and sporadic reports estimate the number of Eritrean immigrants in the United States to be around 150, 000. While at home and through their journey to the US, these immigrants have had multiple experiences of traumatic events which affect their mental health in the process. Therefore, studying mental health conditions like depression among this vulnerable group of immigrants is crucial. However, a culturally appropriate measurement scale to study prevalence of depressive symptoms is lacking. This study aims to translate and adapt the existing CES-D scale into Tigrigna language for use in Eritrean immigrants in the United States. The development of this tool will help to detect depressive symptoms in the Eritrean immigrant population and potentially also Tigrigna speaking Ethiopian immigrants. Detection of such symptoms will serve not only to facilitate early referral and treatment but also can be used for planning and designing preventive measures.

<center>**Chapter Two**</center>

<center>**Review of Literature**</center>

Introduction

This Chapter provides an extensive review of the literature related to depression

in immigrant populations and appropriate methods to measure depression in this specific

population. A search of articles published in English using terms: Eritrea, immigrant,

refugee, mental health, depression, and United States in CINAHL, Pub Med, PsycINFO,

Web of knowledge, and Google scholar data bases could not retrieve even a single article

on Eritrean immigrants. Then, the search was broadened to African immigrants using the

combination of key words mentioned above. Closer review of pulled articles using this

search method showed that most studies were done based on immigrants from diverse

parts of the world and in most cases the description of immigrants' background is not

clear. Therefore, the author decided to include all relevant studies irrespective of the

origin of the immigrants included in the study.

Eritrean Immigrants in the United States: Historical Perspectives

Before the 1970s, migration of Eritreans was rare. The first mass immigration

started in the mid and late 70s when the war between the Eritrean People Liberation Front

(EPLF) and Ethiopia reached its climax. The problem continued until Eritrea gained its

independence in 1991. As a result, in 1991 about one-third of the total Eritrean

population was in exile. Between 1991 and 1998; war, destruction, and oppression seem

<center>11</center>

to be replaced with a bright future with promises of development and prosperity for Eritrea. However, in 1998 another border conflict erupted between Eritrea and Ethiopia. The war combined with political suppression and endless military conscription created new waves of emigration. After spending many years of hardship in refugee camps, Eritreans who met immigration criteria of countries like the United States were granted asylum status and resettled. At present, the total number of Eritrean immigrants in the United States is estimated to be more than 150, 000 (Kibreab, 2009). According to recent report by the United Nations Higher Commission for Refugees (UNHCR), Eritrean immigrants receive the highest approval rate of asylum all over the world (UNHCR, 2009).

From 1971 to 1980, a total of 2,991 African refugees were admitted into the United States. Many of these individuals (1,307 or 43%) came from Ethiopia. During this time Eritrea was considered to be part of Ethiopia and immigrants from Eritrea are included in this report as Ethiopians. The number of refugees admitted from Africa increased 10 fold to 22,149 in the 1980s. In 1991-1992 alone, almost 10,000 African refugees were admitted as permanent residents, with the majority of these individuals (6,850 or 74%) from Ethiopia. This surge of African immigrants to the United States underscores the importance of educating health care workers about the special medical needs of this group (Getahun, 2007).

Background Information on Depression

Depression is one of the oldest known psychiatric problems. Yet, in spite of the enormous scientific work done in the field, depression remains among the major health problems worldwide affecting about 121 million people. According to the World Health

Organization (2008) report, depression is the leading cause of disability as measured by Years Lived with Disability (YLDs) and the fourth leading contributor to the global burden of disease Disability Adjusted Life Years (DALYs) in 2000. The same report projected that by the year 2020, depression will be second in the ranking of DALYs calculated for all age groups and both genders. Currently, depression is second in the DALYs ranking for males and females 15 to 44 years of age. In the United States major depressive disorder affects approximately 14.8 million (6.7%) in a given year (WHO, nd). There are no statistics on the prevalence of depression in Eritrea or Eritrean immigrants in the United States.

The annual cost of depression in the year 2000 was estimated to be around 83.1 billion US dollars in lost workdays each year. Of this total cost, 31% was associated with direct treatment cost, 7% was related to post depression suicide, and the remaining about 55% was associated with work place costs (Greenberg et al., 2003). Two-thirds of the 30, 000 reported suicides in the US each year are related to depression. According to the 2004 US mortality statistics, suicide was the 11th leading cause of mortality (Center for Disease Control and Prevention, 2005). Anecdotal reports from different media sources and personal contacts show that there are incidences of suicide among Eritrean immigrants. However, there is no evidence if depression is a factor.

Mental Health and Immigrants

There is minimal research in the area of mental health among African immigrants to the United States. Available literature about mental health issues of African immigrants is mainly from European countries, Canada, and Australia. Most of these studies reported increased rates of mental illness among African immigrants, compared

with native born individuals. For example Selten & Sijiben (1994), reported higher initial admission rates for schizophrenia among young male immigrants from Morocco compared with their native born counterparts. Severe psychopathology has also been reported among Ethiopian immigrants in Israel (Arieli & Ayche, 1993). In a study done in Sweden, immigrants reported poorer mental health than the Sweden-born population (Sundquist, Behmen-Vincevic, & Johansson, 1998; Persson, 2000).

In contrast, several studies have reported better health in immigrants' compared to the endogenous society (Chen, Ng, & Wilkins, 1996; Collins & Shay, 1994; Hernandez & Charney, 1998; and Vega et al., 1998). These findings are not consistent with the mental health and migration model (Beiser, 1990). This model explains how premigration and post migration experiences cause stress and how other variables such as age, gender, social support, and other personal attributes affect mental health outcomes of immigrants. Findings of the above studies could be the result of the effectiveness of the immigrant selection process and may require further investigation. For example, Fenta, Hyman, and Noh (2004), studied a random sample of 342 Ethiopian immigrants and refugees in Toronto and found a higher rate (9.8%) of depression in Ethiopian immigrants compared to the general public in Ontario (7.3%). In this study depression was measured using the Composite International Diagnostic Interview questionnaire

It is evident that significant gaps exist in knowledge of mental health needs of African immigrants in the United States. For example there were no studies on in the health status of Eritrean immigrants in the United States. These gaps can be addressed by developing culturally appropriate measurement tools to collect appropriate data to understand mental health care needs of immigrants from Eritrea.

Types of Immigrants

The terms refugees and immigrants have been used interchangeably in the literature. Although a refugee movement is similar to immigrants' movement, the former is used for individuals who left their territory because of political events rather than economic interest. In this study the term immigrants is used to describe all Eritreans who left Eritrea during the long war for independence with Ethiopia and those who continue to flee from militarism, political repression, fear of another round of boarder conflict with Ethiopia, and those who came through programs like US Diversity Visa lottery program, family reunion, or education, and currently reside in the United States.

Overall immigrants are grouped into three different categories: voluntary, involuntary, and illegal/undocumented. Voluntary immigration indicates that the individual's migration process had element of his/her personal choice. According to different reports, most immigration is voluntary, but personality characteristics, combined with family dynamics, may diminish the voluntary nature of the decision to resettle particularly in the case of the young or seniors. Involuntary immigration is often a common characteristic of refugees, whose freedom of choice is limited or taken away by local, coercive social forces. Illegal immigrants are those who bypass official channels for residency. This third group is believed to be at greater risk of depression compared to the former two groups of immigrants because of their unique source of stress from fear of detection by officials in the host country.

Depression in Immigrants

For decades it has been said that depressive symptoms are common in immigrant populations mainly during the first few years of resettlement. Such conclusion has been widely accepted because there is a general understanding that the immigration process is characterized by hardships and challenges that are believed to negatively impacts immigrants' mental health. A number of studies reported that rates of mental health problems are relatively higher among migrants (Chung, Bemak, Ortiz, & Sandoval-Peres, 2008).

Some of the experiences faced by immigrants are loss/death of friends or family members, torture, and unhealthy living environment (Bemak, Chung, & Pedersen, 2003). Finding employment with a reasonable salary is also a challenge mainly because of the language barrier or certification issues within host country and could lead to depression (Hermansson, Timpka, & Thyberg, 2002). The settlement process for most immigrants includes experiences like loss of status, employment, family, possession, absence of social support, discrimination, and significant cultural differences between the country of origin and location of resettlement (Bhugra, 2003). Researchers who investigated the impact of immigration on mental health do not include a strong argument for migration specific effects. Mental health problems in immigrants could be a result of multiple factors like general health including mental health, employment status, and immigration itself (Wren & Boyle, 2001).

Epidemiology of depression

Epidemiologic evidence of prevalence of mental health problems, including depression in immigrants ranges widely and/ or conflicts with each other making conclusions difficult to made (Hollifield et. al., 2002). One of the biggest issues when it comes to determining the epidemiology of depression is the issue of the conceptualization and measurement of the term depression by researchers. Cross culturally comparable ways of describing the nature and experience of well-being and illness are obviously needed to study mental disorders in different ethno-cultural groups. In part, the language used by a specific group to explain health status is influenced by what group members consider relevant to their feelings, experience, and symptoms (Aday, Ronald, & Gretchen, 1980; Jones & Korchin 1982). This cultural selectivity in talking about mental health conditions include beliefs about illness and the symbolic meanings attached to mental disorders. Therefore, a detailed understanding of the nature of mental disorders in specific ethno-cultural groups is very important to design a valid measurement scale for that specific group. While translating already existing validated tool into another language is appropriate, researchers should pay extra attention beyond word for word translation because usage of words to express feelings of mental states may not be the same in the original and target language. If concerns of conceptualization, measurement tool, and methodology are not addressed, measuring immigration specific depression will be difficult.

Center for Epidemiologic Studies Depression (CES-D) scale

The Center of Epidemiologic Studies Depression scale (Radloff, 1977) has been one of the most commonly used depression scales. A Pub Med search using the key

17

phrase 'Center for Epidemiologic Studies Depression Scale' retrieved 2592 articles. It was

developed with an intention to have a short and cost effective depression tool for use in

community surveys (Comstock & Helsing, 1976, Weissman et al., 1977) by the Center

for Epidemiologic Studies at the National Institute of Mental Health (NIMH).The

development of the CES-D scale was not guided by a theoretical framework. Rather is

was developed by incorporating items from previously developed depressive inventories,

including Zung's Depression Scale (Zung, 1965), Beck's depression Inventory (Beck et

al., 1961), a scale developed by Raskin et al. (1967), and a scale developed by the

Minnesota Multiphase Personality Inventory (MMPI).

   The CES-D scale measures the breadth of depressive disorder symptoms. It is a

screening tool for use in community based surveys. It cannot be used to measure progress

of depressive symptoms. Therefore, it cannot be used to assess the effectiveness of

treatment or intervention. It takes about five minutes to complete the 20 item CES-D

scale tool. Items were selected from other tools based on their relevance to six major

symptom areas indentified (mood, guilt, hopelessness, psychomotor retardation, loss of

appetite, and sleep disturbance). The 20 items used in this tool are therefore believed to

fit into one of these six areas of depressive symptoms. It is one of the most accepted and

frequently used tools to measure symptoms of depression (Murphy, 2002). The CESD

scale is cost effective and can be administered, scored, and interpreted by a lay person.

The scale is brief and fits a single A-4 size paper making its administration easier.

Each item is measured on a 4 point Likert scale that ranges from zero to three. A value of

zero for an item indicates that the symptom occurred rarely, i.e., less than one day or not

occurred at all and the maximum value 3 means the item under investigation occurred

most or all of the time, i.e., five to seven days. For sixteen of the twenty items in the instrument, a higher score represents more impairment. For the remaining four items, the cores are reversed, i.e., higher score represent less impairment. The total score can range from zero to sixty and higher scores are interpreted as more depressive symptoms. A score of 16 or more is associated with the presence of depressive disorders irrespective of socio-demographic variations. The CES-D scale uses ordinal categories that express the breadth of symptoms of depressive disorder for the purpose of epidemiologic screening.

Measurement equivalence of CES-D scale

The CES-D scale has been widely accepted for community based epidemiologic studies. However, group differences are reported in the CES-D items across diverse racial/ethnic groups (Cole et al., 2000; Perreira et al., 2005), age (Gatz & Hurwicz, 1990; Hays, Landerman, George, Flint, Koening, Land et al., 1998), gender (Posner, Stewart, Marin, & Perez-Stable, 2001), and instrument language (Roberts et al., 1990). These observed group differences in the CES-D scale could be true or artificial differences. The effect of race/ethnicity on the measurement properties of the CES-D has not been fully identified. Most previous studies of measurement bias in the CES-D scale have focused on subscale instead of individual item analysis to trace differential responses across racial/ethnic groups (Nguyen et al., 2004; Perreira et al., 2005). Cole and colleagues (2000) compared two racial/ethnic groups and found evidence for racial/ethnic item bias in the CES-D scale. The authors suggested that Blacks were more likely to endorse higher levels of the two interpersonal problem items ("people are unfriendly" and "people disliked me").

Cross-cultural Applicability of CES-D Scale

The accuracy with which the CES-D can assess depressive symptoms across racially or ethnically diverse groups of people is one of the critical questions in cross-cultural depression research. Because optimal depression screens and optimal cut-scores have not been identified for racially or ethnically diverse populations like immigrants, it is important to review evidence for and against the utility of the CES-D scale for depression in a group of immigrants and also to identify if the existing optimal cut-score works also in different immigrant populations. Given the fact that numerous cross-cultural and cross-national studies on depressive symptoms have used the CES-D scale and the main purpose of this proposal is to translate, adapt, and then evaluate the measurement properties of the CES-D scale in Eritrean immigrants and refugees here in the United States, it is wise to review studies that addressed the cross-cultural applicability of the CES-D scale.

The CES-D has been used in a number of cross-cultural populations to measure depressive symptoms by scholars in various fields of study. The majority of these studies have compared prevalence rates and means of the CES-D across racial/ethnic groups and reported evidence of differences in both prevalence rates and means across those groups (Foley et al., 2002; Mackinnon, McCallum, Andrews, & Anderson, 1998). For example, comparing group means of the CES-D scale among Japanese, Taiwanese, African Americans and Whites in the United States., Krause and Liang (1993) showed that Japanese elders had the lowest mean scores on overall depressive symptoms, followed by Taiwanese, Whites, and African Americans. In a study by Inoba and colleagues (2005) Japanese also demonstrated lower mean scores on the CES-D scale than Whites.

Some studies reported prevalence of depressive symptoms as measured by the CES-D scale across and within racial/ethnic groups to be 3.5% for Germans (Papassotiropoulos & Heun, 1999); 13.2% for Hispanics and 9.2% for Whites (Swenson et al., 2000); 14% of African Americans (Foley et al., 2002); 19.8% for African Americans (Baker, Velli, Freidman, & Wiley,1995); 25.3% for Koreans (Cho, Nam, & Suh, 1998); 25.4% for Mexican Americans (Gonzalez, Haan, & Hinton., 2001); and more than 30% for Korean Americans (Jang Kim & Chiriboga, 2005). These studies used the cut of point of 16 as recommended by the initial developer. Given the evidence for different means and rates of probable depression across racial/ethnic groups, a major issue with regard to cross-cultural applicability of the CES-D instrument is the extent to which such racial/ethnic group comparisons reflect true differences in the depressive symptoms or conversely, how much is due to measurement variance in the construct of interest.

Factor analyses of the CES-D have been conducted since its initial development. In the first study, four specific factors were identified (Radloff, 1977). These four factors are described as depressed affect (DA), positive affect (PA), somatic vegetative (SV), and interpersonal (IP). Several studies have found that the CES-D has acceptable internal consistency as well as Radloff's (1977) four-factor solution of depressive symptoms in different racial/ethnic groups (Blazer, Landerman, Hays, Simonsick, & Saunders, 1998; Krause & Liang, 1993; Roberts, 1980). Two studies by Roberts and colleagues (Roberts, 1980; Roberts, Vernon, & Rhoades, 1989), for example, showed acceptable reliability of the CES-D in Mexican-Americans, African Americans, and Anglo Americans. However, subsequent factor analysis suggested different factors than four factors-suggested by the original author. Some examples of studies that used factor analysis to

indentify factor structures of CES-D and confirm or fail to confirm existing structures include: a two factor structure of depressed affect (DA) & positive affect (PA) (Edman et al., 1999); a three factor structure of DA and somatic, positive/wellbeing, and interpersonal(IP) (Guarnaccia, Angel, & Worobey, 1989); a four factor structure of DA, somatic-retarded activity, PA/well-being, and IP(Golding, Aneshensek, & Hough,1991; Radloff, 1977; Shafer, 2006); and a seven factor structure of DA, Somatic-retarded activity, PA/well-being, IP, anxiety, introspection and crying (Callahan & Wolinsky, 1994).

The inconsistency in the factor structure of CES-D in different racial/ethnic groups and patient groups is a cause of concern that cultural differences might exist in the way individuals express depressive symptomatology (Rhee, 1999). Variations in samples and methodology used in the studies above make it hard to compare findings. But, cross-cultural studies using the CES-D scale support its general usefulness for assessing depression in diverse groups of adults. Diverse findings from studies mentioned above might suggest that cultural factors may impact the way people from different racial or cultural background report depressive symptoms.

<div align="center">

Health Disparities

</div>

Cross cultural studies in the past few years reported significant physical health and mental health differences across different ethnic/racial groups (Neighbors, Trierweiler, Ford, & Muroff, 2003). However, it is not well known whether these observed differences across culture are true differences, or just reflect cultural bias in the measures used. The assumption that concepts can be measured in the same way for all groups of people with different race/ethnic and cultural background is not always

realistic. For example, an item might not have the same meaning for either raters/interviewers or respondents of different ethnic/racial backgrounds. Such difference in interpretation of the meaning of a given item will impact the way the individual will report about his/her health. True extent of a disorder among groups might be masked if measures that do not take into account the cultural norms of a particular group are used. This might not be an issue in variables that do not rely on self-report. However, for many other constructs that affect health outcomes, such as depression, conceptual variations in self-reported measurements among different cultural groups exist and this will affect accuracy of measurement.

Measurement Equivalence in Cross-cultural Study

One major issue related to assessing the cross-cultural comparability of depressive symptoms has been the equivalence of measures (Bravo, 2003; Liang, 2002). The issue of equivalence is more important when self-report screening measures are involved (Liang, 2002). A valid and reliable measurement scale in one language often loses meaning and context after translation. Even with accurate translation, the problem of different nuances unique to different cultures may not be completely avoided (Bravo, 2003). Failure to substantiate the equivalence in a depression instrument in different languages for different group of people is potentially serious because it may lead to inaccurate prevalence rate and misleading group comparisons (Vandenberg & Lance, 2000).

Conceptual Equivalence

Conceptual equivalence is the most basic type of equivalence and implies that research materials or observed behaviors have the same meaning in two or more cultures (Liang, 2002). Hui and Triandis (1985) identified conceptual equivalence as a necessary

condition for making cross-cultural comparisons. Conceptual equivalence can be evaluated by using back-translation, focus groups, and in-depth interviews.

Metric Equivalence

Assuming conceptual equivalence, metric equivalence assures that a given measurement specification can be applied to different cultures (Liang, 2002). Metric equivalence occurs when the factor loadings of items in the depression instruments are invariant across two or more cultural groups (Crockett, Randall, Shen, Russel, & Driscoll 2005). Confirmatory factor analysis (CFA) is a widely used approach to evaluate metric and structural equivalence simultaneously (MacCallum & Austin, 2000).

Structural Equivalence

Structural equivalence refers to similarities in the causal mechanism between a construct of depression and its consequences across different racial/ethnic groups (Liang, 2002). Structural Equation Modeling (SEM) and path analysis have been widely recommended to evaluate structural equivalence (Crockett et al., 2005). Most researchers agree that SEM is the most versatile approach to evaluating metric and structural equivalence simultaneously (Byrne & Watkins, 2003; Liang, 2002; MacCallum & Austin, 2000).

Approaches for Cross Cultural Studies of Depression

Cross-cultural comparative studies involve consideration of two central perspectives called the emic and etic approaches (Brislin, Lonner, & Thorndike, 1973; and Canino, Lewis-Fernandez, & Bravo, 1997). The former approach exhibits the internal perspective of the researcher, while the later exhibits the external perspective of the researcher (Morris, Leung, Ames, & Lickel, 1999). Emic approach Utilizes

24

characteristics and observations that are culturally unique to a particular group at a given period in time (Rait & Burns, 1998). This approach does not support comparative (e.g. cross-cultural) research as it looks at variables in terms of language and culture. Therefore, instrument developed in one culture may not be useful in other culture. The etic approach, on the other hand, is comparative in nature and is directed at extracting standardized categories of phenomena excluding local distinctions (Canino et al., 1997).

Emic Approach

The emic approach explores the internal logic of a given culture and its uniqueness because the approach considers these factors as a necessary step prior to any valid cross-cultural analysis. Hence, it does not allow cross-cultural comparisons using identical constructs and standardized diagnostic interviews as case-finding instruments (Cheng, 2001). The emic approach is quite useful in understanding the relatively unique characteristics of the manifestation of depressive symptoms in a given ethnic group because it focuses on examining a construct from perspectives of a specific culture. That is, understanding that construct of interest as individuals from within that culture understand it (Schaffer & Riordan, 2003).

Etic Approach

The etic cross-cultural research approach on depression presumes that the etiology of depression is global and key constructs of depression exist equally across the boundaries of all cultures. Often times cross-cultural researchers utilize the etic approach for the cross-cultural comparative study of depression highlighting the search for equivalence across cultures and using similar methods, constructs, and measures across groups in an attempt to increase the generalizability of their research findings (Schaffer &

Riordan, 2003; van de Vijver, 2001). In sum, both emic and etic approaches have their own advantages and disadvantages. Hence, combining the two approaches would benefit cross cultural researchers who are interested in translation and adaptation of measurement scales.

Translation of Measurement Scales into another Language

Translation of measurement scales from one language to another has been practiced for decades. The Council of International Test Commissions (ITC) recognized the importance of developing guidelines for test translation and adaptation in 1991.This resulted in the development of the first guidelines in 1994 which were updated in the year 2005 to meet the even growing field of psychometrics (Schaffer & Riordan, 2003).

Translation Methods

Using a previously developed measurement scale with established psychometric properties for cross cultural research has been a common practice. This practice has been advocated for its cost effectiveness and its contribution to psychometric properties of the scale used in cross cultural studies (Yu, Lee, & Woo, 2004). Translation of a scale from source language to a target langue requires more than proficiency in the source and target languages. It requires rigorous methodology and in-depth understanding of the construct under consideration from the perspectives of both cultures (John, Hirsch, Reiber, & Dworkin, 2006).

The recommended procedure for translating research instruments is known as the back translation method. The objective of using this method is to ensure the equivalent meaning of items in both languages. This is accomplished by having questions in the source language translated by bilingual person, preferably from the target culture, into the

target language. Another bilingual individual then translates the items from the target language back to the source language. The two source language versions are then compared for equivalence. This process can be repeated until satisfactory equivalence is obtained.

The shortcoming of this procedure is that researchers have no knowledge of the number of translators required to accomplish the task, because the procedure recommends translation and back translation until concept equivalency is reached between the original instrument and the translated version of the instrument. Pilot testing of translated scale in a selected sample from the target population is essential. Pilot testing can help to identify and fix problems which were not identified by translators.

Challenges of Translations

Brislin and Colleagues (1973) identified three major potential errors committed by researchers during the process of translation. The first most common error is distortion of the meaning of a word or phrase. Second, adding extra meaning to the intent of the original item, and thirdly deleting meaning of some words in the original version inappropriately. Often, these errors are detected during the back translation process.

<div align="center">

CES-D Scale Translations

</div>

The CES-D scale has been translated into different languages for use including: Spanish (Roberts et al., 1989; Chiriboga, Jang, Banks, & Kim, 2007), Japanese (Shima, Shikano, Kitamura, & Asai, 1985; Kanazawa, White, & Hampson, 2007), Chinese (Cheung & Bagley, 1998), Italiana (Fava, 1983), Greece (Fountoulakis et. al., 2001), Arabic (Ghubash, Daradkeh, Naseri, Al Boushi, & Daheri, 2000), Armenian (Van Trtan, 1997), Portuguese (Goncalves & Fagulha, 2004), Korean (Noh, Avison, & Kaspar 1992),

Germany (Hautzinger, 1988), French (Fuhrer & Rouillon, 1989), Indian languages

(Gupta, Punetha, & Diwan, 2006) etc. The English version of CES-D is available for use

for free and can be requested from the National Institutes of Epidemiology Branch.

It is not known if the CES-D has been translated into any language for use in an

African population. Journals published in most African countries might not be included

in widely used data bases such as those used to search articles for this review. The

translation of the CES-D scale into Tigrigna language would help to test the cross cultural

use of the instrument in immigrant population from Africa. Translated instruments need

to undergo vigorous tests of measurement invariance testing subsequent models in an

orderly and systematic manner. Judgment about the best model that fits the data is made

based on multiple fit indices. Below is a summary of some of the most commonly used fit

indices in the literature.

Absolute Fit Indices

Absolute fit indices establish how well a model fits the sample data (McDonald

and Ho, 2002) and ranks models based on the superiority in fit. These measures are used

as basic guidelines of how well the proposed theory fits the data. Their advantage over

the incremental fit indices is that, they do not require baseline model to compare with.

They measure how well the model fits the data without the need of comparison (Jöreskog

and Sörbom, 1993). Examples of such indices include Chi-Squared test, RMSEA, GFI,

AGFI, the RMR and the SRMR.

Chi-Square (χ^2)

The Chi-Square measures overall model fit and it gauges the extent of

inconsistency between the sample and fitted covariances matrices' (Kline, 2010). An

adequate model fit would give an insignificant result at a 0.05 threshold (Barrett, 2007),

thus the Chi-Square statistic is often referred to as either a 'badness of fit' (Kline, 2010).

One of its limitations is that it assumes multivariate normality and severe deviations from

normality may result in model rejections even when the model is properly specified

(McIntosh, 2006). Another disadvantage of this test is that it depends on sample size

(Jöreskog and Sörbom, 1993). To overcome the second limitation the ratio of chi-square

to the degrees of freedom is used and the acceptable ratio for this statistic,

recommendations range from 2.0 to 5.0 (Tabachnick and Fidell, 2007).

Root mean Square Error of Approximation (RMSEA)

RMSEA is a measure of how well a model would fit the populations' covariance

matrix (Byrne, 1998). Its advantage is that it is sensitive to the number of estimated

parameters in the model. In other words it favors a model with fewer numbers of

parameters. It is believed that an RMSEA of 0.08 to 0.10 provides an average fit and

below 0.08 shows a good fit (MacCallum, Browne, & Sugawara, 1996). However, more

recently, a cut-off value close to .06 (Hu and Bentler, 1999) or a strict upper limit of 0.07

(Steiger, 2007) seems to be acceptable.

Goodness-of-Fit Statistics (GFI) and the Adjusted Goodness-of-Fit Statistics (AGFI)

The Goodness-of-Fit statistic (GFI) was introduced by Jöreskog and Sorbom

(Tabachnick and Fidell, 2007). Its values range from 0 to 1 with larger samples

increasing its value. When there are a large number of degrees of freedom in comparison

to sample size, the GFI demonstrates a downward bias (Sharma, Mukherjee, Kumar, &

Dillon, 2005). The usual cut-off point recommended is 0.90 but some recommended a

higher cut-off point of 0.95 (Miles and Shevlin, 1998). Like the GFI, values for the AGFI

also range between 0 and 1. Overall values of 0.90 or greater indicate well fitting models. They are both sample sensitive.

Root mean Square Residual (RMR) and Standardized Root Mean Square Residual (SRMR)

The RMR and the SRMR are the square root of the difference between the residuals of the sample covariance matrix and the hypothesized covariance model (Kline, 2010. The range of the RMR is calculated based upon the scales of each indicator; therefore, if a questionnaire contains multiple scaling, the interpretation of RMR becomes a problem (Kline, 2010). The standardized RMR (SRMR) addresses this problem. Values for the SRMR range from 0 to 1.0. Models with values less than .05 are considered good fit models (Diamantopoulos & Siguaw, 2000). According to Hu and Bentler (1999), values up to 0.08 are also acceptable.

<div align="center">

Incremental Fit Indices

</div>

Comparative Fit Index (CFI)

The Comparative Fit Index (Bentler, 1990) is a revised form of the NFI which takes into account sample size (Byrne, 1998). The main advantage of this index is that it performs well even when sample size is small (Tabachnick & Fidell, 2007). Initially it was introduced by Bentler (1990). The values for this statistic range between 0.0 and 1.0 with values closer to 1.0 indicating good fit. A value of CFI $\geq$ 0.95 is often times accepted as indicative of good fit (Hu & Bentler, 1999).

<div align="center">

Normed Fit Index (NFI)

</div>

NFI is known to assess a model by comparing the $\chi2$ value of the model to the $\chi2$ of the null model. Values for this statistic range between 0 and 1 with Bentler and Bonnet

(1980) recommending values greater than 0.90 indicating a good fit. The short coming of

NFI is that like the chi-square, it is sensitive to sample size, and tends to under estimate

fit when the sample size is below 200 (Bentler, 1990).

Chapter Three

Methods

Introduction

Chapter three describes the methods used in translating the CES-D scale and analytical steps used in establishing the psychometric properties of the Tigrigna version CES-D scale. Contents include: details of study design, participants and settings, instrument/questionnaire used, translation procedures, qualitative and quantitative data analyses plan, and ethical dimensions of the study. Limitations are reviewed and plan for dissemination of the outcomes of the study is also included.

Study Design

This study used a mixed method design- qualitative and quantitative. The English version 20-item CES-D scale was translated into Tigrigna through forward and backward translation technique. The study has four major phases including forward-back ward translation, cognitive interview, pilot test, and re-test. Each phase is discussed in detail below under procedure section. A secondary data that consists of CES-D scores for men and women 45 to 75 years old who are able to undergo baseline and annual follow-up visits from a prospective cohort study (see Aiyer, et al., 2007) is used to test factorial invariance of the CES-D scale across groups (Tigrigna speaking Eritrean immigrants and English speaking non Eritrean US citizens).

Participants and Settings

The participants of this study are Tigrigna speaking Eritrean immigrants or refugees to the United States, aged 18 to 64 years (M = 37.6), at the time of interview. Immigrants who identify themselves as Eritrean descendants and who claim to be fluent in Tigrigna were recruited to be part of the study. Community gatherings such as church, wedding ceremonies, and other social gatherings were used to recruit eligible participants. Generally, a sample size of 10 participants per item is considered to be adequate for factor analysis (Everitt, 1975). The CES-D scale used in this study has twenty items. Therefore, we planned to recruit between 200 and 250 Eritrean Immigrants or Refugees in the United States and ended up recruiting a total of 260 participants. Participants with missing data for three or more items are dropped (n=7). When the number of missing data is less or equal to two items, the mean score for the individual is substituted (n=18). In addition, ten non Eritrean US citizens and ten Eritrean immigrants or refugees in the United States were recruited using convenient sampling technique to participate in the cognitive interview (second phase of the study). Participants of the cognitive interview were not allowed to take part in the pilot and re-test phases of the study.

In addition to the CES-D score data collected from Tigrigna speaking Eritrean immigrants here in the united states, CES-D score from English speaking American citizens in a study called "The Heart Strategies Concentrating on Risk Evaluation (HEART SCORE) study" [see Aiyer, 2007] is used for the purpose of testing factorial invariance across the two cultural groups. Participants in this study were 1918 English speaking healthy non Eritrean adults who are 45 to 75 years old (M= 59.1) at the time of

interview. Participants in the HEART SCORE study completed the 20-item CES-D scale

as part of a structured interview containing standard socio-demographic items, medical

history, and other scales. The final sample size used for invariance analyses include 253

Eritrean immigrants (male =164, female = 89) and 1918 Americans from the HEART

SCORE study.

Table 1

Demographic Characteristics of Tigrigna and English Samples.

Characteristics	N (%)		Mean(SD)	
	Tigrigna	English	Tigrigna	English
Age in Years				
18 to 24	23 (9.1)			
25 to 34	86 (34)			
35 to 44	85 (33.6)			
45 to 54	33 (13)	580 (30.2)		
55 to 64	26 (10.3)	829 (43.1)		
65 to 75	-	509 (26.7)		
Total			37.5 (10.7)	59.1 (7.4)
Gender				
Male	166 (65.6)	661 (34.5)		
Female	87 (34.4)	1257 (65.5)		

Note: SD = Standard deviation; N = Number; % = Percentage

Instruments/Questionnaire

Instrument. In this study, the English version of the 20-item CES-D scale was

translated into Tigrigna. The CES-D scale is a 20-item self administered scale that

measures depressive symptoms during the week before the date of interview. Each item is

measured on a four point Likert scales that range from zero to three. A value of zero for

an item indicates that the symptom occurred rarely or not occurred at all and the

maximum value 3 means the item under investigation occurred most or all of the time,

i.e., five to seven days. For sixteen of the twenty items in the instrument, a higher score

represent more impairment. For the remaining four items (item 4, 8, 12, & 16), the scores

are reversed, i.e., a higher score represents less impairment. Reponses to the 20-item

CES-D scale are summed to identify the level of depressive symptoms experienced by

34

individuals in the past week. A score of 16 or more indicates depressive risk (Radloff, 1977).

Internal consistency ranging from 0.8 to 0.9 has been reported. Test-retest reliability within two weeks was reported to be within the range of 0.4 and 0.7 (Devins et al., 1988). The scale also discriminated between depressed and non-depressed individuals in a sample of adolescent participants (Dierker et al., 2001). Table 2 lists the items associated with the four factors [depressive affect (DA), positive affect (PA), somatic vegetative (SV), and interpersonal (IP)] of the CES-D scale.

Participants (Eritrean sample) were asked to provide brief demographic information including gender, age, and year of immigration. The Institutional Review Board (IRB) of the University of South Florida (USF) approved this study. The secondary data used is from a study that was approved by respective IRB. Therefore, no additional consent was required for the study.

Table 2

CES-D Scale Items, Abbreviations, and Factors.

Item Content	Abbreviation	Factor	Item
I felt that I could not shake off the blues even with help from my family or friends.	Blues	DA	3
I felt depressed.	Depressed	DA	6
I thought my life had been a failure.	Failure	DA	9
I felt fearful.	Fearful	DA	10
I felt lonely.	Lonely	DA	14
I had crying spells.	Cry	DA	17
I felt sad.	Sad	DA	18
I felt that I was just as good as other people.	Good	PA	4
I felt hopeful about the future.	Hopeful	PA	8
I was happy.	Happy	PA	12
I enjoyed life.	Enjoy	PA	16
I was bothered by things that usually don't bother me.	Bothered	SV	1
I did not feel like eating; my appetite was poor.	Appetite	SV	2
I had trouble keeping my mind on what I was doing.	Mind	SV	5
I felt that everything I did was an effort.	Effort	SV	7
My sleep was restless.	Sleep	SV	11
I talked less than usual.	Talked less	SV	13
I felt that people disliked me.	Get going	SV	20
People were unfriendly.	Unfriendly	IP	15
I felt that people disliked me.	Disliked	IP	19

DA= Depressed Affect, PA= Positive Affect, SV= Somatic/Vegetative, IP= Interpersonal

Procedures

Forward and backward translation. Forward and backward translation is a

well-known method that ensures semantic and content equivalence between the original

and translated version (Behling & Law 2000). The principal investigator identified four

bilingual competent translators. Translators were provided with adequate instructions by

the principal investigator regarding inferences, wording, and phrasing. The importance of

rigorous trans-cultural translation was emphasized. The first two bilingual translators

(TP_1 & TP_2) translated the CES-D scale from the original English version to the target

Tigrigna independently. Then the two translators (TP_1 & TP_2) discussed and agreed upon

the initial translated version to be used for back translation. The other two bilingual

translators (BTP$_1$ & BTP$_2$) then independently back translated the CES-D scale from the target Tigrigna version to the original (English) version and then discussed and agreed upon the back translation. The four bilingual translators (TP$_1$, TP$_2$, BTP$_1$, & BTP$_2$) then discussed and agreed upon the draft translation of the CES-D scale. Then the two versions of the CES-D scale (Tigrigna and English) were compared for content and semantic equivalence. This was done by three raters who examined the original scale, the Tigrigna version, and the back translated version. The three raters scored each item in the Tigrigna version in a seven point scale ranging from perfectly equivalent (7) to not at all equivalent (0). The scores from the three raters were averaged for each item and items with a score of 4 or less were translated and back translated again by separate bilingual individuals and another round of rating by the same three raters until an average of 4 point score is achieved by each item. Then the Tigrigna version of the CES-D scale became ready for use in cognitive interviews.

   Cognitive interview. Cognitive interviews were developed to make sure that survey questions are interpreted in the manner they were intended. This strategy is accepted as a valuable part of evaluation in the process of developing questionnaires (Jobe & Mingay, 1989; Alaimo, Olson, & Frongillo, 1999; Carbone, Cambell, & Honess-Morreale, 2002). In this study, before using the newly translated Tigrigna version CES-D scale for the pilot test, cognitive interviews were done by the principal investigator to assess equivalence of comprehension between the Tigrigna and English version of the 20-item CES-D scales. Participants in the cognitive interview were asked to read or listen to each item in the 20-item CES-D scale and then paraphrase their understanding. Ten Tigrigna speaking Eritrean immigrants were asked to explain their understanding of the

meaning of specific expressions or words that the investigator identified as possibly

causing difficulties in comprehension.

Concerns about any item (even if expressed by one person) were considered as a

problem and participants were asked to identify alternative wording to help clarify the

item or phrasing of the instrument's instructions to better convey the intended meaning.

Subsequent interviews used the same translated Tigrigna version of the CES-D scale per

the protocol for cognitive interview. When a participant encountered a problem that had

also been identified by a previously interviewed participant, the alternative wording

suggested by the previous participant were used to determine if it helps to clarify the

intended meaning of the question or wording of the instructions. Figure 1 reviews the

steps used to translate the CES-D scale and subsequent analyses plans.

English Version
(CES-D scale) ──────────────────────────────────→ Cognitive Interview
                                                    (US citizens)

Phase I

    Two translators(TP1&TP2)
    Translate independently to
    Target language

        Phase II TP1 & TP2 sit together &
                  Agree on preliminary translation

            Phase III

                Two translators (BTP1 & BTP2)
                Back translate to source language

                    Phase IV   Final Version of
                                                       Tigrigna CES-D
                        BTP1 & BTP2 sit &   Scale
                        Agree on Preliminary BT
                                                                Pilot test & Re-test

                        Phase V TP1&TP2 discuss With BTP1
                                 &BTP2 & agree on translation
                                                       Test Psychometric
                                                         Properties
                            Phase VI 3 raters rated each item On 7pt. Scale.
                                      Items less or equal to 4, Back to Phase I

                                Draft translation ──────→ Cognitive Interview
                                                           (Eritrean Immigrant/Refugee)
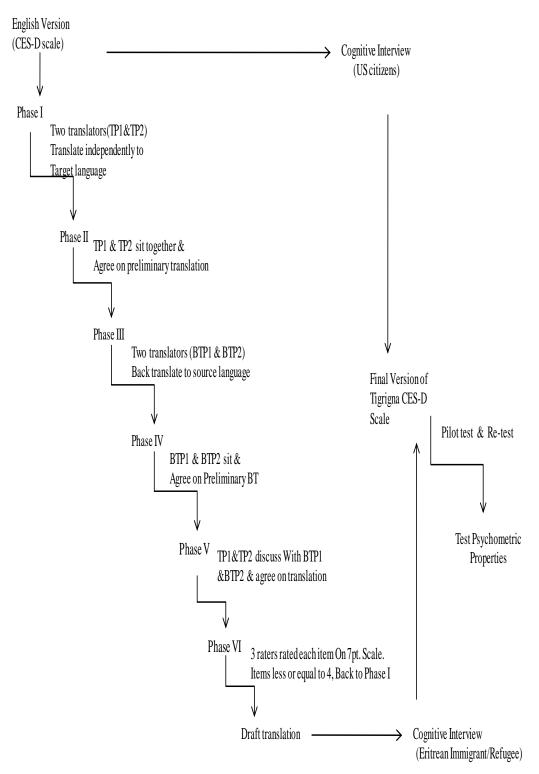
Fig *1. Steps for Translation of CES-D Scale from English to Tigrigna*

Testing translated Tigrigna version of CES-D scale. Information about the study was disseminated by distributing pamphlets at Eritrean community gatherings. Church and community leaders were also contacted to help with disseminating information about the study. Potential adult participants (18 to 64 years old) were asked to complete the Tigrigna version of CES-D scale (see appendix A) and brief demographic information and return the packet to the researcher using a self addressed, stamped envelope. In the case of data collection in places where community event (church, weddings, etc) took place, participants were asked to complete and return the completed questionnaire to a drop box. Participants of this study were those who identify themselves as immigrant or refugee from Eritrea and able to read and write Tigrigna.

Re-test. To test the stability of the Tigrigna version CES-D scale, CES-D score was collected from thirty participants who consented to complete the CES-D scale for the second time one week after the initial test.

Statistical Analysis Plan

Qualitative Analysis

The audio taped cognitive interviews were reviewed by the primary investigator. Handwritten notes taken during the interviews were also triangulated to provide additional information and clarity. The data were used to assess the equivalence of comprehension and response error across the two languages. Based on the findings alterations were made to the CES-D scale questionnaire before the pilot testing to increase the quality of data collected. This will give more confidence in the results from comparative analysis of the two data sources.

Preliminary Quantitative Analysis

Descriptive statistical analysis is done to investigate the shape of the data distribution and the pattern of response for each item of the 20 item CES-D scale across the two groups. Reliability analysis is used to asses scale reliability and test-retest reliability of the Tigrigna version CES-D scale. Internal consistency reliability was assessed using SPSS reliability procedure to get information on the correlation between each particular item and the test scale score minus the item, and the proportion of variance in each item that can be explained by the other items.

Analysis of Factor Structure

A confirmatory factor analysis using a statistical package LISREL 8.72 (Joreskog & Sorbom, 2001) is done to determine a statistically acceptable fit of the data collected using the Tigrigna version of the CES-D scale to the originally suggested four factor structure (Radloff, 1977). In this confirmatory factor analysis we hypothesized the following a priori: (a) The CES-D scale is explained by four underlying factors (SV, DA, PA, and PI); (b) Each item would have a nonzero loadings on the underlying factor it was designed to measure and zero loadings on all other factors; (c) the four factors would be inter-correlated; and (d) the error terms associated with each observed item (i.e., the uniqueness) would be uncorrelated. Figure 2 below presents schematic presentation of the hypothesized model.
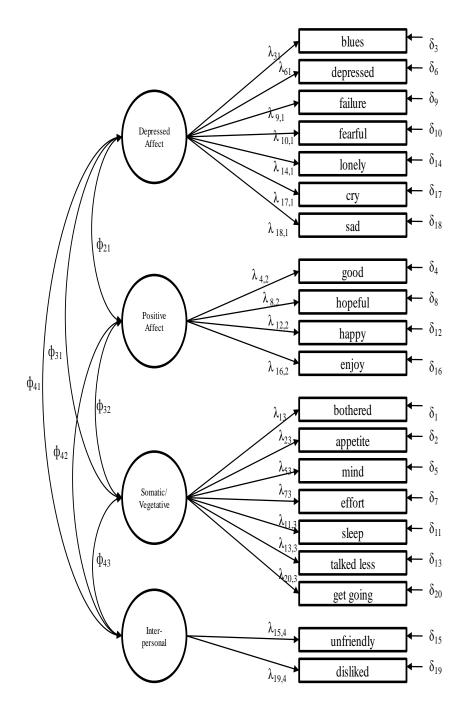
Figure 2. *Hypothesized model of factorial structure of the 20-item CES-D scale (Radloff, 1977).*

Testing Measurement Invariance

Translation adequacy will be demonstrated using the invariance testing procedure. This will be done by comparing the findings from the Tigrigna version with findings of secondary data from "The Heart Strategies Concentrating on Risk Evaluation (HEART SCORE) study" (see Aiyer, 2007) which used the twenty item CES-D scale in healthy American citizens, 45 to 75years of age. Item loadings, reproducibility of the originally recommended 4 factors, and cross cultural invariance of the CES-D items was investigated using systematic step by step approach recommended in structural equation modeling literature.

We considered two typical methods of testing adequacy of an estimated structural model reported in the literature. The maximum likelihood (ML) provides a likelihood ratio chi-square statistic to test whether the moment matrices reproduced from the estimated parameters differ significantly from the observed sample moment matrices. A single chi-square value assessing aggregate fit across the estimated and observed matrices of two groups can be obtained in situations where two groups or stacked modeling is used (as in the case of this study). A statistically significant chi-square renders rejection of the tested model in favor of an alternative model. On the other hand, if the Chi-square value is statistically non-significant, the tested model will be considered to be adequate representation of the data.

In addition to investigating to the Chi-square values for each model, the Chi-square values for stacked structural models were also assessed. Subsequent models were generated by placing restrictions on parameter estimates systematically and the difference in chi-square values for the two nested models, i.e. the $\Delta\chi^2$, and degrees of freedom (the

difference in degrees of freedom for the two models) under consideration, or Δdf are used to test whether the model fits the data or not. Using $\Delta\chi^2$ value, we tested the statistical significance of the difference in fit of subsequent increasingly restrictive models. When the $\Delta\chi^2$ value is statistically significant, the less restrictive model provides a significantly better fit to the data and vice- versa. In an effort to test the cross cultural invariance of the 20-items of the CES-D scale in the two samples, we used the general guideline proposed by Vandenberg and Lance (2000).
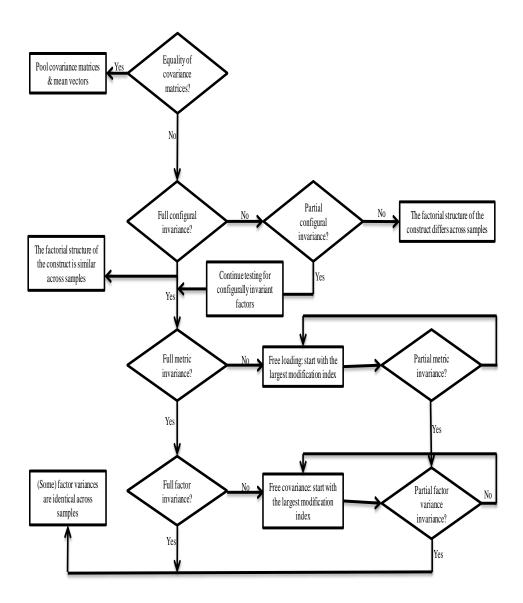
Figure 3. *Flowchart of the logical sequencing for assessing cross-group invariance (adapted from Vandenberg & Lance, 2000, p.56)*

Note: In this process we were not interested in comparing means across the two samples and scalar invariance test was not included in the flowchart. The analysis proceeds from assessing metric invariance to assessing factor invariance. The following multi-group confirmatory factor analyses are done based on the suggested guidelines above

Model zero (configural invariance). This model is the first and basic form of measurement invariance and it is tested by specifying the same measurement model for both Eritrean immigrants and non Eritrean us citizens. Both the number of factors and the factor-indicator correspondence are the same, but all parameters are allowed to be freely estimated within each sample. Configural invariance is said to be satisfied if the hypothesized four factor model structure fits across the two samples, suggesting that individuals from both samples conceptualize the construct depression in the same way. In this study we tested configural invariance by running individual CFA in each group then by running Multi-Group Confirmatory Analysis (MGCFA) by constraining the factorial structure to be same across groups. This model is used as a standard for comparison of subsequent models and hence is termed baseline model.

Model one (metric invariance). This is a stronger form of invariance and it is tested by (a) constraining all factor loadings to be the same across the two samples and (b) comparing the two hierarchical models, one that corresponds to constrained factor loading hypothesis and the other corresponds to the configural invariance hypothesis using the chi-square difference test ($\Delta\chi^2$). If the hypothesis that the un-standardized factor loadings of each CES-D item are equal across the two samples is retained, then we can conclude that the construct depression is manifested the same way in each group. If the $\Delta\chi^2$ for the comparison stated above is not statistically significant, then the fit of the

model with the equality constrained factor loadings is not significantly worse than that of

the model without these constraints, hence the hypothesis of equal factor loadings across

two samples is retained. In this case, we can test stronger form of measurement

invariance. However, if this hypothesis is rejected, the less strict hypothesis of partial

measurement invariance is tested by releasing one cross-group equality constraint at a

time based on the size of their modification indices. This procedure will help us to

identify sources of stress that are responsible for metric non-invariance (Cheung &

Rensvold, 2002). On subsequent tests, we let factor loadings of these items be freely

estimated in each sample while the loadings of the remaining indicators are constrained

to be equal across both samples. According to Vandenberg & Lance (2000), at least

partial metric invariance is a pre-requisite for subsequent tests of invariance. Therefore, if

the data show metric invariance, we will continue to the next level of invariance testing.

Model two (full factor invariance). This model is tested by constraining all

factor covariances and variance's to be the same across samples. Consistent with the

strategy we followed in testing previous models, model two is tested by specifying a

model in which the factor loadings from model 2 remain invariant and all factor

covariance's and variance's are constrained to be equal. To test the hypothesis that the

structural relations among the four factors of the CES-D scale are equivalent across the

two samples, we compare the fit of model 2 with that of model 1.

In order to evaluate structural models, we used multiple fit indices. Widman and

Reise (1997) in an effort to make model fit evaluation reliable, inclusive, and acceptable,

recommended the following two basic strategies (a) Reporting of two or more fit indices

for each model i.e., according to these authors use of multiple fit indices will allow

47

researchers to demonstrate the same characteristics of model fit using multiple indices, (b) Incorporate theory and professional judgment with statistical analyses to achieve a reasonable decision of accepting or rejecting a given model.

With emphasis to the above two suggestions, we used multiple indices to test models to our data. Fit indices used in this study include: (a) Chi-square; (b) the comparative fit index (CFI) by Bentler (1990); (c) the root-mean-square error of approximation (RMSEA) proposed by Steiger and Lind (1980); (e) the goodness-of-fit index (GFI) and (f) the standardized root-mean-square (SRMR). Details on fit indices mentioned above and additional fit indices are discussed in chapter two of this paper.

Summary

Maximizing the validity of inferences to be made about group similarities and/ or differences is one of the main goals of cross-cultural studies. Achieving such goal requires strong design and analyses plan. Especially cross-cultural studies that involve translation and adaptation of scales need to take into account linguistic, cultural, and psychometric considerations in every step of the process. The use of appropriate analytical procedure then helps to maximally eliminate bias and achieve acceptable level of equivalence. The methods used in this study are designed with the goal of translating and adapting the 20-item CES-D scale in to Tigrigna language to develop a scale which is free from bias. The rigorous methods used in this study are believed to lead to the development of a Tigrigna CES-D scale that can be used to make defensible inferences about cross-cultural similarities and differences observed in depression symptom scores in Tigrigna speaking Eritrean immigrants compared to other groups.

<center>**Chapter Four**</center>

<center>**Results**</center>

Introduction

    Chapter Four presents the findings of this study. The chapter begins with

preliminary analyses that show the characteristics of the data used in the analyses and the

response pattern of respondents. Reliability analysis and invariances tests including

configural, metric, and factor variance-covariance invariances are presented.

Preliminary Analyses

    Equivalence in distributions of the items. The equivalence in data distribution

of the items was investigated by determining the shape of each item distribution.

Descriptive statistics and frequencies were done using SPSS 19. The findings from these

analyses are summarized in Table 3 and Table 4 below. The scores for each item ranged

from 0 to 3 in both samples indicating that the same scoring patterns were used. Table 3

summarizes the values of skewness and kurtosis across the two groups. Ten items had

skewness that approached zero in the Tigrigna data as compared to only one item with

skewness that approached zero in the English CES-D data. Sixteen of the 20 items in the

Tigrigna CES-D scale had kurtosis values close to zero compared to only one value close

to zero in the English CES-D scale. According to Kline (2010), absolute values greater

than 10.0 for kurtosis index show some problem with the data and values greater than

<center>49</center>

20.0 are considered extreme. Chou & Bentler (1995) suggests a value of 3.0 for skewness

index as extreme.

Table 3.

Skewness and Kurtosis of the Twenty CES-D items across Tigrigna and English samples.

Item	Tigrigna Sample		English Sample	
	Skewness	Kurtosis	Skewness	Kurtosis
Bother	0.86	- 0.13	1.55	1.74
Appetite	1.16	0.19	2.95	9.07
Blues	1.39	0.72	2.30	5.15
Good	0.23	- 1.57	2.70	7.57
Mind	0.90	- 0.33	1.34	1.43
Depressed	1.53	1.38	1.72	2.72
Effort	0.75	- 0.59	1.66	2.27
Hopeful	0.78	- 0.80	2.75	7.91
Failure	1.32	0.63	3.20	10.90
Fearful	1.17	0.25	2.18	5.00
Restless	0.80	- 0.28	0.77	- 0.36
Happy	0.73	- 0.56	1.59	2.43
Talked	0.90	- 0.05	2.29	5.31
Lonely	1.36	0.74	1.97	3.72
Unfriendly	1.35	0.69	2.79	8.60
Enjoy	0.63	- 0.81	2.71	7.74
Cry	1.57	1.57	3.66	14.76
Sad	1.36	0.91	1.72	3.00
Disliked	1.67	2.13	3.39	13.13
Get going	0.84	- 0.27	1.45	1.90

After evaluating the distributions of the data, how individuals from the two

groups responded to each item in the CES-D scale was assessed. Findings in Table 4

demonstrate a difference in patterns of item response especially in categories of moderate

amount of the time and most of the time. The percentage of individuals who endorsed

these two categories is smaller in the English CES-D group.

Table 4

Pattern of Item Responses: Tigrigna and English Samples.

	Responses (%)			
Items	None of the time	A Little of the time	A Moderate amount of the time	Most of the time
Bother				
Tigrigna	41.5	37.2	13.0	8.3
English	65.3	23.6	8.2	2.9
Appetite				
Tigrigna	57.7	22.5	12.6	7.1
English	85.6	10.7	3.1	0.6
Blues				
Tigrigna	64.8	17.4	11.1	6.7
English	77.5	16.3	4.4	1.8
Good				
Tigrigna	38.7	17.4	17.0	26.9
English	81.6	13.7	3.2	1.5
Mind				
Tigrigna	51.8	25.3	17.4	5.5
English	58.4	32.0	7.2	2.3
Depress				
Tigrigna	67.6	18.6	10.7	3.2
English	67.5	24.9	5.5	2.1
Effort				
Tigrigna	43.9	29.6	16.6	9.9
English	65.4	25.0	6.4	3.2
Hopeful				
Tigrigna	49.0	22.5	14.2	14.2
English	82.4	13.1	3.2	1.3
Failure				
Tigrigna	61.3	21.3	11.1	6.3
English	86.2	10.1	2.7	1.0
Fearful				
Tigrigna	59.7	21.3	14.2	4.7
English	78.0	18.6	2.9	0.6
Restless				
Tigrigna	41.5	35.6	15.0	7.9
English	40.1	36.1	15.3	8.5
Happy				
Tigrigna	38.7	35.2	14.2	11.9
English	63.5	29.3	5.0	2.2
Talked				
Tigrigna	45.8	34.0	14.2	5.9
English	78.4	16.7	3.9	1.0
Lonely				
Tigrigna	65.2	18.2	13.0	3.6
English	72.3	21.0	4.8	1.9

Table 4 (continued)

Pattern of Item Responses: Tigrigna & English Samples

	Responses (%)			
Items	None of the time	A Little of the time	A Moderate amount of time	Most of the time
Unfriendly				
Tigrigna	59.7	23.7	7.9	8.7
English	85.2	13.0	1.6	0.3
Enjoy				
Tigrigna	39.1	31.2	16.6	13.0
English	81.8	13.9	3.0	1.4
Cry				
Tigrigna	66.0	20.6	8.7	4.7
English	89.8	8.0	1.8	0.3
Sad				
Tigrigna	62.1	22.9	11.1	4.0
English	67.6	26.6	4.2	1.6
Dislike				
Tigrigna	66.4	22.5	7.1	4.0
English	87.3	10.4	1.7	0.7
Get going				
Tigrigna	46.6	31.6	16.6	5.1
English	61.5	30.5	6.0	2.0

Reliability of the CES-D Scale

Table 5 shows the item-total correlations of the 20 items of the CES-D scale for both samples. Cronbach's alpha of the whole scale also shows adequate reliability of the CES-D scale in both samples. Test re-test in one week period of time in the Eritrean immigrants sample showed a correlation of 0.91 suggesting stability of the scale. The values are consistent with findings reported in the literature.

Table 5.
Corrected Item-Total Correlation and Cronbach's Alpha

	Correlated total-item correlation	
Item	Tigrigna	English
Bother	.39	.24
Appetite	.35	-.23
Blues	.54	.39
Good	.18	.38
Mind	.53	.30
Depress	.65	.48
Effort	.51	.29
Hopeful	.14	.40
Failure	.50	.48
Fearful	.57	.66
Restless	.58	.57
Happy	.35	.32
Talked	.55	.76
Lonely	.59	.58
Unfriendly	.44	.61
Enjoy	.40	.26
Cry	.56	.54
Sad	.53	.63
Dislike	.51	.63
Get going	.49	.30
Cronbach's alpha	.86	.91

Testing Invariance across Groups

In testing for the invariance of the CES-D scale across the two samples, the following three hypotheses were considered: (a) the four factors underlying depression in the CES-D scale are equivalent; (b) the pattern of factor loadings is equivalent across the two samples; and (c) the structural relations among the four factors of depression are equivalent.

Multiple models were tested to look for equivalencies of the CES-D items across the two samples using a set of standard parameters to guide subsequent tests in a logically ordered and increasingly restrictive style. The findings of these tests are summarized in Table 6. In order to test equivalence of a given scale across groups, establishing a well-fitting baseline model for each group separately is a prerequisite. Therefore; the original

four factor structure of the CES-D scale (Radloff, 1977) for each sample separately was tested. The baseline four factor CES-D scale model originally suggested for the general population is presented in Figure 2. This model fitted the data from both samples well as evidenced by multiple fit indices. The fit indices for the Tigrigna sample were ($\chi^2 =$ 299.87, df = 164, RMSEA = .06, SRMR = .06, GFI = .89, and CFI = .98) and for the English sample ($\chi^2 =$ 1496.81, df = 164, RMSEA = .07, SRMR = .04, GFI = .92, and CFI = .98).

In this single group confirmatory factor analysis, almost all of the fit indices were very similar across the two groups suggesting that the correlated four factor model was supported by both samples. In order to cross validate the correlated four factor model of the CES-D scale, multi-group confirmatory factor analysis (MGCFA) was done to further test configural invariance.

Configural invariance. The single group CFA tested whether each sample adequately fits the correlated four factor model separately. Since the two samples separately demonstrated good fit to the correlated four factor model, generally it was expected that the MGCFA model (M_0 in Table 6) also would show adequate fit. As expected the MGCFA showed reasonably adequate fit ($\chi^2_{(328)} =$ 1796.68, RMSEA= .07, SRMR = .06, GFI = .89, CFI = .98). It is important to note that the χ^2 value for model M_0 is simply the sum of the χ^2 for the two separate models ('a' and 'b' in Table 6) in the single group confirmatory factor analysis. These findings suggest that the configuration of factor indicator relationships of the CES-D scale was equivalent across the two independent samples. However, this test doesn't give enough information on the invariant

operation of each item in the CES-D scale and the factor loadings. Therefore, additional test of metric invariance was required to address the issue.

Metric invariance. An analysis was conducted to examine how the content of each CES-D item was being perceived and interpreted across the two samples by running the next model M_1 (full metric invariance). Having established a good fit baseline (configural: M_0) model, the hypothesis of metric invariance was tested by constraining the matrix of factor loadings ($\Lambda_{Tigrigna} = \Lambda_{English}$) to be equal across the two samples. This hypothesis was tested by modifying M_0 i.e. putting the constraint on the Λ matrix and we identified this model as M_1. The χ^2 shown in Table 6 from M_1 was compared with χ^2 from M_0 because the metric invariance model was stacked within the configural model. Findings of this analysis showed that the restricted, stacked model resulted in a significant $\Delta\chi^2$ when compared to the less restricted model. Therefore, the hypothesis of full metric invariance was not tenable. To identify the sources of stress or misfit within this model, a detailed evaluation of the modification indices (MIs) provided by the LISREL output was completed. Initial review showed that item 17 (cry) had the largest source of stress. Thus this item's loading was allowed to vary across groups and the model was re-evaluated. The fit indices showed slight improvement but still the model was statistically different from the baseline model suggesting that the constraints specified in the more restrictive model (M_1) do not hold (i.e., M_0 and M_1a are not equivalent across groups). The process of freeing items with greater source of stress continued systematically until a model which was not different from the baseline model was achieved (i.e., the $\Delta\chi^2$ value when compared with the baseline model was statistically non-significant to suggest the tenability of all specified equality constraints).

To achieve this, in addition to item 17 (cry); items 10 (fearful), 19 (disliked), 13 (talked less), 2 (appetite) and 11 (sleep) were freed sequentially. After freely estimating 6 of the 20 items in the CES-D scale, partial metric invariance was achieved [(M_1f- see Table 7) χ^2 (df) = 1811.95(338), $\Delta\chi^2$ (Δdf) = 15.27(10), RMSEA = .07, SRMR = .07, GFI = .89, and CFI = .98)]. The differences in fit indexes for subsequent alternative models are reported in Table 7. Note that, since at least one factor loading per latent variable was constrained to be invariant in the baseline model, the full metric invariance model was tested by constraining the remaining 16 factor loadings to invariance across the two samples.

Full Factor invariance. According to Beckstead, Yang, and Lengacher (2008); even in the presence of variation in factor loadings across two samples (i.e., partial metric invariance) there is a possibility of having equivalent factor variances and covariances across samples. Byrne et al. (1989) suggested that a test of factor variance and covariance can be done provided there is at least one metric invariant item in each factor. Therefore, since the data satisfied this criterion we proceeded with the next phase of analysis. To test the invariance of factor variance-covariance matrices across the two samples in this study, constraints were imposed on all elements of the phi matrix to be invariant across the two samples and named this model M_2 (see Table 6). The comparison between M_2 and M_1f showed a statistically significant change in fit. The SRMR = .12 was poor making the hypothesis of full factor invariance untenable. The principle of freeing parameter estimates with greater sources of stress or misfit using the size of modification indices as a criterion was followed. To reach to the final model (M_2g: see Table 7), 7 of the 10 constraints we initially imposed on the phi matrix were freed. The bottom section

of Table 8 summarizes those results. Interesting findings that must be noted include: the correlation between PA and DA was smaller for the Tigrigna sample than the English speaking sample. The same was true for correlations between PA and SV, although correlations between IP and DA and IP and SV were lower for the English sample than the Tigrigna sample.

Table 6.
Summary of model fit indices for the correlated four factor CES-D scale

Model	χ^2(df)	$\Delta\chi^2$ (Δdf)	RMSEA	SRMR	GFI	CFI
Single group CFA models						
a. Tigrigna Sample	299.87(164)		.06	.06	.89	.97
b. English Sample	1496.81(164)		.07	.04	.92	.98
Multi-group CFA models						
M_0: Configural model	1796.68(328)		.07	.06	.89	.98
M_1: Full metric model	1895.84(344)	99.16(16)	.07	.09	.86	. 98
M_1f: Partial metric model	1811.95(338)	15.27(10)	.07	.06	.89	.98
M_2: Full factor invariance	2058.63(348)	246.69(10)	.07	.12	.80	.97
M_2g: Partial factor invariance	1815.80(342)	3.85(3)	.07	.09	.89	.98

χ^2 = Chi Square, RMSEA= Root Mean Square Residual, CFI=Comparative Fit Index, RMR= Root Mean Square Residual, GFI= Goodness-of-Fit Index.

Table 7.

Detailed Description of Model Fit Statistics for the Correlated Four factor CES-D Scale

Model	χ^2(df)	$\Delta\chi^2$(Δdf)	RMSEA	SRMR	GFI	CFI
Single group CFA						
a. Tigrigna Sample	299.87(164)		.06	.06	.89	.98
b. English Sample	1496.81(164)		.07	.04	.92	.98
Multi-group CFA						
M_0: Configural Invariance	1796.68	328	.07	.06	.89	.98
M_1. Full metric Invariance	1895.84(344)	99.16(16)	.07	.09	.86	.98
M_1a. Item 17 FR	1861.67(343)	64.99(15)	.07	.09	.87	.98
M_1b. Item 10 FR	1839.72(342)	43.04(14)	.07	.08	.88	.98
M_1c. Item 19 FR	1832.08(341)	35.41(13)	.07	.08	.88	.98
M_1d. item 13 FR	1825.22(340)	28.54(12)	.07	.08	.89	.98
M_1e. item 2 FR	1818.54(339)	21.86(11)	.07	.07	.89	.98
M_1f. Partial metric Invariance	1811.95(338)	15.27(10)	.07	.07	.89	.98
M_2: Full factor Invariance	2058.63(348)	246.69(10)	.07	.12	.80	.97
M_2a. ɸ22 FR	1878.84(347)	66.89 (9)	.07	.12	.86	.98
M_2b. ɸ21 FR	1868.42(346)	56.47 (8)	.07	.12	.87	.98
M_2c. ɸ32 FR	1858.11(345)	46.16 (7)	.07	.11	.87	.98
M_2d. ɸ42 FR	1844.38(344)	32.43 (6)	.07	.11	.88	.98
M_2e. ɸ43 FR	1841.66(343)	29.71 (5)	.07	.10	.88	.98
M_2f. ɸ41 FR	1835.13(342)	23.19 (4)	.07	.10	.88	.98
M_2g. partial factor Invariance	1815.80(342)	3.85 (3)	.07	.09	.89	.98

χ^2 = Chi Square, RMSEA= Root Mean Square Residual, SRMR= Standardized Root Mean Square Residual, GFI= Goodness-of-Fit Index, CFI= Comparative Fit Index.

Table 8.
Common Metric Standardized Structural Coefficients for the Correlated Four Factor
CES-D Scale Items in Tigrigna and English Samples.

Item	Brief description	Tigrigna Sample				English Sample			
		DA	PA	SV	IP	DA	PA	SV	IP
3	Blues	.51	-	-	-	.51	-	-	-
6	Depressed	.52	-	-	-	.60	-	-	-
9	Failure	.44	-	-	-	.36	-	-	-
10*	Fearful	.53	-	-	-	.32	-	-	-
14	Lonely	.52	-	-	-	.46	-	-	-
17*	Cry	.48	-	-	-	.22	-	-	-
18	Sad	.47	-	-	-	.52	-	-	-
4	Good	-	.50	-	-	-	.37	-	-
8	Hopeful	-	.46	-	-	-	.44	-	-
12	Happy	-	.60	-	-	-	.60	-	-
16	Enjoy	-	.46	-	-	-	.42	-	-
1	Bothered	-	-	.44	-	-	-	.45	-
2*	Appetite	-	-	.39	-	-	-	.20	-
5	Mind	-	-	.56	-	-	-	.47	-
7	Effort	-	-	.57	-	-	-	.57	-
11*	Sleep	-	-	.62	-	-	-	.44	-
13*	Talked less	-	-	.55	-	-	-	.35	-
20	Get going	-	-	.50	-	-	-	.46	-
15	Unfriendly	-	-	-	.30	-	-	-	.30
19*	Disliked	-	-	-	.28	-	-	-	.45

Φ Matrices

	Tigrigna Sample					English Sample			
	DA	PA	SV	IP		DA	PA	SV	IP
DA	1.13				DA	.98			
PA	.26*	1.36*			PA	.99	.94		
SV	.91	.26*	1.01		SV	.87	.87	1.00	
IP	.82*	.18*	.82*	1.88*	IP	.60	.63	.54	.82

Note: DA = depressive affect, PA = positive affect, SV = somatic vegetative, IP = interpersonal. All coefficients are presented in standardized form and are significant (P< .05). Factor loadings are given from column 3-10. Off diagonal elements of the φ matrices are correlations among factors; diagonal elements are factor standard deviations.
*Coefficients were non-invariant across the two samples (p < .05).

Table 9.

Non-Equivalent Parameters across Tigrigna and English Speaking Groups

Parameter	Item Content	Related Factor
Factor loadings		
Item 2	My appetite was poor	Somatic vegetative
Item 10	I felt fearful	Depressed affect
Item 11	My sleep was restless	Somatic vegetative
Item 13	I talked less than usual	Somatic vegetative
Item 17	I had crying spell	Depressed affect
Item 19	I felt that people disliked me	Interpersonal
Variances & Covariances		
PA & PA		Positive affect/Positive affect
PA & DA		Positive affect/Depressed affect
SV & PA		Somatic vegetative/Positive affect
IP & PA		Interpersonal/Positive affect
IP & SV		Interpersonal/Somatic vegetative
IP & DA		Interpersonal/Depressed affect
IP & IP		Interpersonal/Interpersonal

Note: PA= positive affect; DA = depressed affect; SV = somatic vegetative; & IP = interpersonal

The six items that were non-equivalent across the Tigrigna and English samples in the initial invariance test analyses are listed in Table 9. Item non-equivalence for these six items could be explained by multiple reasons, such as (a) poor translation, (b) sample difference in demographic characteristics that are relevant to the topic under investigation e.g., age, gender, and (c) lack of appropriateness of the item content for the Tigrigna sample. To address (a), the mean of the translation scores given by the three raters were compared using t-test. Findings of this t-test showed that there was no statistically significant difference between the non-equivalent and equivalent items ($t = -.18$, $df = 7$, $p = .86$).

To address (b), an attempt was made to match the two samples on variables such as sample size, gender, and age. Sample size and gender distribution were perfectly matched to the Tigrigna sample in a sub sample of English speakers. However, because of the initially wide age variation among the two samples, our attempt to match the groups in terms of age was not fully successful with the mean being 37.5 years in the

Tigrigna sample and 50 years in the English sample. Invariance tests were conducted to
see if the non-invariant items from the initial analysis would remain non-invariant in the
matched sample. Test for configural invariance showed that the data fit the correlated
four factors model of the CES-D scale adequately (see Table 10 for fit indices). Test for
metric invariance demonstrated partial metric invariance. However, in this analysis we
detected only three non-invariant items (Item 17, Item 10, and Item 19). All three items
identified to be non-invariant are the first three of the six non-invariant items in the initial
analysis. The fit indices of increasingly restrictive models tested using closely matched
Tigrigna and English samples are given in Table 10 below and the findings are discussed
in detail in Chapter 5.

Table 10.
Detailed description of model fit statistics for the correlated four factor CES-D after
matching the Tigrigna and English samples.

Model	χ^2(df)	$\Delta\chi^2$(Δdf)	RMSEA	SRMR	GFI	CFI
Single group CFA						
a. Tigrigna Sample	299.87(164)		.06	.06	.89	.98
b. English Sample	420(164)		.08	.05	.85	.97
Multi-group CFA						
M_0: Configural invariance	720.10(328)		.07	.06	.89	.97
M_1. Full metric invariance	820.22(344)	100.12(16)	.08	.10	.87	.96
M_1a. Item 17 FR	763.05(343)	42.95(15)	.07	.08	.88	.96
M_1b. Item 10 FR	747.21(342)	27.11(14)	.07	.07	.89	.97
M_1c. Item 19 FR	739.01(341)	18.91(13)	.07	.07	.89	.97

χ^2 = Chi Square, RMSEA= Root Mean Square Residual, SRMR= Standardized Root Mean Square
Residual, GFI= Goodness-of-Fit Index, CFI= Comparative Fit Index.

Table 11.

Depressive symptom scores across Tigrigna and English samples stratified by age and gender.

Characteristics	N (%)			
	Tigrigna Sample		English Sample	
	≥ 16	< 16	≥ 16	< 16
Overall risk of depression	113(44.7)	140(55.3)	237(12.4)	1681(87.6)
Risk of depression by age				
18 to 24	8 (7.1)	15 (10.7)		
25 to 34	44 (38.9)	40 (28.6)		
35 to 44	34 (30.1)	50 (35.7)		
45 to 54	17 (15)	15 (10.7)	96 (40.5)	484 (28.8)
55 to 64	10 (8.8)	20 (14.3)	114 (48.1)	715 (42.5)
65 to 75	-	-	27 (11.4)	482 (28.7)
Risk of Depression by gender				
Male	68 (60.2)	98 (70)	60 (25.3)	601 (35.8)
Female	45 (39.8)	42 (30)	177 (74.7)	1080 (64.2)

Chapter Five

Discussion

Introduction

This chapter begins with emphasis on the importance of measurement invariance in cross-cultural research. Then the findings of the study are discussed and limitations are also presented. Finally conclusions, implication for nursing, and future research directions are discussed.

Over the past ten years, the CES-D scale has been examined for its cross cultural measurement equivalence for multiple racial and ethnic groups (e.g., Cole et al., 2000; Crockett et al., 2005; Jang et al., 2005; Kim, Chiriboga, & Jang, 2009). Such efforts emphasize the increasing recognition of and importance of measurement equivalence in cross-cultural research. Whether observed variation in psychometric test scores is attributed to an actual difference in construct that a given instrument or scale measures is an important question to ask in many research domains. For example it is vital for any cross-cultural researcher to demonstrate evidence of measurement invariance when working on cross-cultural study where the same scale is used to measure and compare the same construct but in two culturally diverse samples. It is essential to demonstrate that the same attribute is related to the same set of observations in the same way for all groups in the study. That is, the relationship between the latent variables and the observations

must be the same in each of the groups selected in order to make meaningful comparisons.

Data analysis in cross-cultural research requires a detailed procedure that begins with careful data screening for detection of problems related to the data. Some data related problems might result in non-positive definite data matrices and others could be the results of violating assumptions like normal distribution of data. In this study, tests for skew and kurtosis showed that items 9 (failure), 17 (cry), and 19 (disliked) were slightly positively skewed and leptokurtic in the English sample (see Table 3), i.e., most of the scores for these three items were below the mean in the English sample. Since the three skewed items were also leptokurtic, procedures like transformation that fix skewed distributions might also fix kurtosis. However, since skew and kurtosis in the English sample are mild or just marginally high, a transformation procedure was not required. Cross-cultural studies assume uniform psychometric properties of the measurement scale used in all samples. To assess these properties at an instrument level, reliability coefficients were assessed in both groups. The internal consistency of the Tigrigna CES-D was ($n = 253$, $\alpha = .86$) and that of the English sample was ($n = 1918$, $\alpha = .91$). These values were high and comparable to the correlations of .84 and .85 reported by Radloff (1977). While the test re-test in one week period of time in the Tigrigna sample showed high correlation ($n = 30$, $r = .91$) suggesting the stability of the scale. Table 5 provides the item-total correlation of the 20 items of the CES-D scale in both the Tigrigna and English sample. Items that do not measure the same construct across the two samples can be identified by comparing the item-total correlations across the two samples, as illustrated in the comparison between the Tigrigna and English sample in Table 5. A comparison of

item-total correlation of the Tigrigna sample and English sample shows that most (13 of the 20 items) item-total correlations were higher in the Tigrigna sample. Such variations might have multiple sources such as: method of administration of the scale, cross-cultural differences, and difference in response style. Both the Tigrigna and English CES-D scales were administered in a pencil and paper form. However, the English CES-D scale was administered as part of broader survey including detailed demographic questionnaires and multiple measurement scales. Hence, participants in the English sample may have experienced greater instrument burden than the Tigrigna sample. Most participants in the Tigrigna sample completed the CES-D scale anonymously on their own (self administered) as opposed to face to face interview in the English sample. It has been suggested that self administered surveys provide more response anonymity to participants' thereby encouraging disclosure of accurate information even in sensitive behaviors related to issues like mental health. For example Chan, Orlando, Ghosh-Dastidar, Duan, and Sherbourne (2004) found an estimated 13% increase in the rates of probable depression when self administered data were used to measure depressive symptoms using the 20-item CES-D scale. Therefore, the significantly high scores in depressive symptoms in the Tigrigna sample compared to the English sample could partly be attributed to the anonymous nature of the administration of the CES-D scale. Table 11 summarizes depressive symptom scores across the two samples stratified by age and gender.

Cross-cultural studies on depressive symptoms are based on data collected from groups from diverse cultural backgrounds. For example in this study the two data sets collected to study depressive symptoms contain responses from people with different

cultural backgrounds and it is possible that the two samples have different response styles. Such difference could be a source of distortion and misinterpretation of findings of the study. Table 4 summarizes the pattern of item responses in the Tigrigna and English samples. In both samples a majority of participants endorsed the first option i.e. "none of the time" more frequently than the other three options. However, close observation of data in Table 4 shows an interesting pattern in the distribution of participant's pattern of item response across the two samples. Except for item 11 ("my sleep was restless"), the percentage of participants who endorsed the "most of the time" option was higher for the Tigrigna sample than the English sample. On the other hand, the percentage of participants who endorsed the "none of the time" option was greater in the English sample except for item 6 ("I felt depressed) and item 11 ("my sleep was restless"). Considering the background of participants in the Tigrigna sample, such variation in response pattern was expected. A majority of participants in the Tigrigna sample had witnessed or suffered multiple traumatic experiences while at their home country or through their course of immigration to the United States. There is a strong relationship between traumatic life experiences and development of post traumatic stress disorder (PTSD) and other mental health disorders (Perkonigg, Kessler, Storz, & Wittchen, 2000; Shalev, et al., 1998). Hence, mental health conditions like PTSD might have contaminated the depressive symptom scores in the Tigrigna sample resulting in higher depressive symptom scores.

To my knowledge, there are no documented statistics on the prevalence of chronic health conditions like hypertension, diabetes, and other cardiovascular disorders in Eritrean immigrants in the United States. However, based on personal observation and

informal conversations with members of the community, it is evident that such chronic health problems are not rare in the community. Even though participants in the Tigrigna sample are drawn from the general population of Eritrean immigrants in the US, presence of chronic health conditions was not ruled out. This might raise a concern about 'criterion contamination' i.e., for example items intended to measure somatic symptoms in the CES-D scale might reflected physical symptoms related to other co-morbidities contaminating the total CES-D score.

This study explored the cross-cultural factorial invariance of the 20-item CES-D scale across Tigrigna speaking Eritrean immigrants/refugees and English speaking US citizens. A correlated four factor model of the CES-D scale consisting of depressed affect (DA), positive affect (PA), somatic vegetative (SV), and interpersonal (IP) dimensions, confirmed an adequate fit for both Tigrigna speaking Eritrean immigrants/refugees and English speaking US citizens i.e., the originally hypothesized correlated four factor model of the 20-item CES-D scale (see Figure 2) was supported both by the Tigrigna and English samples. As described in chapter four of this paper, configural invariance was established suggesting the adequacy of the hypothesized model across both samples. In other words, all items of the CES-D scale loaded on their respective factor in both samples.

The importance of multi-group confirmatory factor analysis (MGCFA) extends far beyond establishing configural invariance (Dimitrov, 2010). It involves other tests of equivalence that involve the specification of cross-group equality constraints for particular parameters. In metric invariance, the loading estimates in Λ were constrained to be equal across groups. The results of this study showed the absence of full metric

invariance. In the absence of full metric invariance, Byrne et al. (1989) recommended one of the following three options: (a) identify reasons for the cross-group differences in factor loadings, (b) remove the non-invariant items (this option assumes that the non-invariant items are few and their removal does not underrepresent given domain), and (c) proceed with partial invariance. In this study, the analytic procedure used involved testing for partial measurement invariance using a stepwise systematic procedure as discussed in the method section of this paper. This procedure lead to partial metric invariance (i.e., factor loadings of 14 of the 20 items in the CES-D scale became invariant across the two samples).

Similarly to test variance-covariance invariance, all elements of Φ were constrained to be equal across the two samples. Full variance-covariance invariance was rejected and further tests resulted in partial variance-covariance invariance (i.e., 3 of the 10 Φ elements became invariant across the two samples). Steenkamp and Baumgartner (1998) suggested that valid cross-national comparisons are warranted so long as at least partial measurement invariance has been demonstrated.

The development of partial measurement invariant models using the procedure used in this paper attracts two important questions (Schmitt & Kuljanin, 2008). The first and most important question is how can one theoretically explain the group differences in factor loadings of the non-invariant items? In other words, can the researcher explain the mathematical differences in factor loadings in terms of existing knowledge and theory about the groups under comparison? For instance in this study six items (appetite, fearful, sleep, talked less, cry, and disliked) were identified as non-invariant items. That is, the six items behaved differently in the two samples under investigation. Review of the factor

loadings of these six items (see Table 8) shows that five of these items (appetite, fearful, sleep, talked less, and cry) have relatively higher loadings in the Tigrigna sample than the English sample. While item-19 (disliked) showed relatively higher loading in the English sample. Then the question is: how do we interpret or explain these differences? Before discussing possible interpretations, the next paragraph will introduce the second question. Then two hypotheses competing with the hypothesis that cultural differences explain the non-invariant items are addressed and method used to rule out them are described.

The second question is how should we handle the non-invariant items? According to Poortinga (1989); the decision about the fate of non-invariant items can vary depending on the interpretation given to them. Strategies to handle non-invariant items generally include: (a) dropping of non invariant items from the scale. However, such a step should be considered only after cautious investigation or when a given item demonstrated characteristics of non-invariance on multiple well designed studies. It is also important to note that dropping an item is possible only when a scale has large number of items to begin with (e.g. Janssens, Brett, & Smith, 1995); (b) the second alternative way of handling non-invariant items is to keep non-invariant items in the model by allowing their loadings to vary while constraining the loadings of invariant items to be equal across groups. This procedure is used in situations where full invariance is unachievable; and (c) the third option to deal with non-invariant items is to use them as indicators of actual difference between groups under comparison. For example in this study, we can use the six non-invariant items, to ask the question why these six items are non-invariant. Is it because of a problem with the translation of these items from the original language? Is it because of the variation in important demographic characteristics across the two

samples? Is there any cultural or other cross-group difference that might explain the situation?

If the answer to the former question is a yes, the solution would be to go back and work on the translation process. While in the case of the latter question, an affirmative answer shows evidence of true cross cultural difference. It is clear that the final outcome solely depends on the accurate identification of non-invariant items. Therefore the use of an appropriate method to test factorial invariance is critical. Hence, before speculating potential explanations for the invariant items and correlations across the two samples, it is essential to rule out the two competing hypotheses (problems related to the translation process and difference in vital demographic characteristics as possible explanations for the non invariant items) discussed above.

As discussed in chapter three of this paper, rigorous procedure was followed in the translation process. A translation and adaptation review form by Hambleton and Zenisky (2011) was used as a guideline by the three raters to evaluate the adequacy of translation and adaptation of the 20-items in the CES-D scale. All items had an average score of 4.3 or more in a 7 point scale. The translation adequacy for the non-invariant and invariant items was also shown empirically to be uniform as evidenced by non-significant t-test (see Chapter 4). In addition, findings of cognitive interviews also did not show evidence of difference in understanding or interpretation of the 20-items of the CES-D scale across the two groups. Therefore, it is unlikely for the findings of this study to be contaminated by issues related to the translation and adaptation process. Thus, the first competing hypothesis of a translational problem is ruled out. The next paragraph

discusses the second competing hypothesis which is related to the variation in important

demographic characteristics (e.g., age and gender) across the two samples.

It has been consistently reported in the literature that the prevalence of depressive

symptoms is higher in women compared to men (Davins and Orme, 1986; Anderson,

Freedland, Clouse, & Lustman, 2001; Katon, et. al., 2004). In the initial analysis of

invariance across the two samples in this study, the proportion of male participants in the

Tigrigna sample was almost two times larger than women participants. While in the

English sample, women participants were almost two times more frequent than male

participants. Participants in the Tigrigna sample were also relatively younger (M = 37.5

years, SD = 10.7, N= 253) than the English sample (M = 59.1 years, SD = 7.4, N= 1918).

To rule out the second competing hypothesis related to these demographic

characteristics, invariance tests were employed after randomly selecting a sub sample of

253 English speakers from the English sample that represented the exact same gender

distribution and more closely matched age (M = 50.1years) with the Tigrigna sample.

Subsequent invariance tests using exact same procedure used in the initial analysis

demonstrated that age and gender didn't seem to affect the findings (see Table 10) of the

study. The metric invariance test detected three items instead of six in the initial analysis.

It is important to note that the three items identified as non invariant in the matched

samples were the same items that were detected as non-invariant in the initial analysis. In

the initial analysis the large sample size increases the power to detect small amount of

non-invariance. Otherwise, the fact that the three non-invariant items detected in the

matched sample were indentified in the same order they appeared in initial analysis

suggests that age and gender didn't matter in the way non-invariant items were detected.

71

It was shown that 6 of the 20 items in the CES-D scale function differently across the two samples in this study. Identifying non-invariant items and providing cautions that these items were different from equivalent samples of individuals from two or more cultures is not enough. Current cross-cultural research requires researchers to move one step more and un-package such cross-cultural differences.

The six non-invariant items in this study include: item 2 (appetite), item 10 (fearful), item 11 (sleep), item 13 (talked less), item 17 (cry), and item 19 (disliked). Table 8 shows that the factor loadings of item 2 (appetite), item 10 (fearful), item 11(sleep), item 13(talked less) were higher in the Tigrigna sample compared to the English sample. On the other hand, the factor loading of item 19 (disliked) was higher in the English sample than the Tigrigna sample. Analysis of invariance using the matched sample resulted in three non-invariant items including item 10 (fearful), item 17 (cry), and item19 (disliked). Scientifically persuasive explanations for cultural differences related to these three items (fearful, cry, and disliked) are provided below.

These findings are not unexpected because the three non-invariant items belong to the PA and IP factors that are known to be very sensitive to social contexts (Noh, Kaspar, and Chen, 1998). According to these authors, the social context of an individual determines the person's values and rules of these interpersonal relationships which in turns dictate his/her perception of relationships with others. The relatively large factor loading of item 19 (disliked) in the English sample is consistent with the literature. In this study, the majority of participants in the English sample were African Americans and it has been consistently reported that African Americans tend to strongly endorse item 19 (disliked) compared to other ethnic groups in the United States (e.g., Cole, et al., 2000).

In Eritrean culture maintaining social agreeability is an important value to become acceptable in society. This might encourage Eritreans to pay extra attention to their relationship with others and hence less likely to endorse item 19 (disliked). Even when they are not happy about their overall relationship with others, Eritreans might not report or under report problems related to their relationship with others in their community. Most Eritrean immigrants have been through extraordinary traumatic life events in their native country and/or throughout their journey to the United States. The task of adapting to a new culture and environment, financial hardship, and uncertainty about family members left behind are also stressful. Generally, to the majority of Eritrean immigrants the journey to the United States was dangerous and filled with uncertainty and fear. Such events make this population susceptible to mental health conditions including PTSD which can cause symptoms like persistent re-experiencing of traumatic events. Therefore, the significantly larger factor loadings for items 10 (fearful) and 17 (cry) in the Tigrigna sample may be explained by possible relationship between past traumatic events and the current state of mental health of this group of population.

In this study, 7 of the 10 ϕ elements were invariant (see Table 8). Before discussing these findings, it is important to address the procedure used to constrain elements of Φ. Decisions to impose constraints were made based on the values of modification indices. For example, to test M_2d, in addition to the constraints made while testing M_1c; ϕ_{42} was constrained because of its relatively larger value compared to the remaining elements of ϕ. However, the change in χ^2 was very minimal and the fit indices also didn't show reasonable improvement from M_2c. Thus, decisions based on

values of the modification indices should be logically and theoretically supported to have a meaningful model that appropriately explains group differences and/or similarities. The Φ matrices given at the bottom part of Table 8 have some interesting findings. For example the correlation between PA and DA is much lower for the Tigrigna sample than the English sample suggesting that the two dimensions are more distinct than they are in the English speaking sample. A similar pattern also applies to the correlation between PA and SV. While in the English sample, the correlation (IP & DA) and (IP & SV) are lower for the English speaking sample than the Tigrigna sample suggesting greater distinctiveness of the respective factors in the English speaking group than in their counterpart Tigrigna speakers. Knowing the way these factors relate to each other could be critical for nursing practice. For example, the correlation between PA & DA is lower for the Tigrigna sample than the English sample. In practice, this would mean that interventions that are designed to improve affect with an attempt to manage depression might not show the same outcome in both groups. Such intervention could be much more appropriate to the English sample than the Tigrigna sample.

Limitations

Limitations of the current study warrant consideration. It should be noted that the findings of this study are confined to the population of Eritrean immigrants/refugees (18 to 64 years of age) from limited parts of the United States. Therefore, findings might not necessarily be generalized to the Eritrean citizens within Eritrea and Eritrean immigrants/refugees across the United States. The comparative analyses in this study used two data sets that were collected using independent research efforts. Hence, variations in study designs and procedures cannot be ruled out as contributing to the

74

results. There were also some procedural differences in the data collection process. In the Tigrigna sample, the CES-D scale was administered with a brief demographic information questionnaire, while the English version CES-D scale was administered as part of a large questionnaire that included other measurement scales and detailed demographic questionnaire. There is a 4 to 6 years difference in when the two data sets were collected. The age and gender distribution was not also proportional in the two data sets. However, analyses with closely matched samples didn't alter the findings of this study. Despite these limitations, this study clearly demonstrated that the Tigrigna version CES-D scale can be used as a screening tool for depressive symptoms in Eritrean immigrants/refugees in the United States. The study also identified items that were non-invariant across the two groups providing baseline for future studies in this population.

Implications for Nursing

The review of the literature on cross-cultural research using the CES-D scale shows that the investigation of measurement invariance is very rare. Demonstration of fit to the original correlated four factor model of the CES-D scale is considered by many as evidence for an invariant scale. The use of such scales by nurses to screen for depressive symptoms could be problematic because of measurement non-invariance among immigrant populations. Therefore, when assessing depressive symptoms across cultures, nurses need to be cognizant of the psychometric properties of the assessment instrument and its cross cultural equivalence.

Nurses in Eritrea can benefit from the availability of this scale in Tigrigna language. While the scale needs further psychometric tests within and outside Eritrea in various representative samples, nurses can start applying the scale for large scale

75

screening of depressive symptoms. Moreover, this work is expected to stimulate discussions and awareness related to cross-cultural assessment among Eritrean nurses and other health professionals.

Various governmental and non-governmental organizations who work with Eritrean immigrants/refugees can also use this scale for screening purposes. Such use may help in the early detection of depressive symptoms in this vulnerable population. Depression is known for its high life time prevalence rate (2–15%) and its substantial association with disability. Unfortunately, according to projections by Murray and Lopez (1997), depression will be the second in terms of its overall burden worldwide. The availability of screening scales like the CES-D will contribute significantly towards the early detection and intervention programs that might alleviate the burden of depression.

Conclusions

Measurement equivalence is important to all comparative studies. It is particularly essential when the comparison being made is cross-cultural because differences in social norms and values clearly affect the way people perceive a construct and the way they perceive the item content. Because of the ever increasing globalization and immigration the need to use measurement scales with individuals from diverse cultural backgrounds is also becoming common practice. This requires translation and adaptation of existing scales to different languages and testing invariance.

Factorial invariance has a critical importance in any cross-cultural research. In the absence of clearly documented factorial invariance, interpretation of differences in score across cultures is meaningless. The assumption of the same conceptualization of items when responding to survey items in not always warranted (Riordan & Vandenberg,

1994). This study is the first step to address the issue of factorial invariance items of the CES-D scale in Tigrigna speaking immigrants/refugees in the United States. Using data collected from Eritrean immigrants/refugees in the United States and a secondary data from the HEART SCORE study (see Aiyer et.al., 2007), evidence of partial measurement and structural invariance of the CES-D scale was shown. The findings of this study provide adequate evidence in support of the applicability of the four factor CES-D scale for measuring depressive symptoms in Tigrigna speaking Eritrean immigrants/refugees. The Tigrigna version CES-D scale has shown psychometric properties comparable to those found in the original population (Radloff, 1977).The next step in the process of developing the Tigrigna version CES-D scale is to conduct multiple studies in various representative samples and establish stronger evidence of its psychometric properties. Studies that compare community samples and clinical samples are also needed to further understand the psychometric properties of the Tigrigna CES-D scale.

Given this is the first study to test invariance of items in the Tigrigna CES-D scale in a representative sample of Eritrean immigrants/refugees, it is important to see if the findings from this study can be replicated. Provided, the non-invariant items detected in this study continue to be non-invariant in future studies based on representative samples of Eritrean immigrants/refugees; additional work to determine the source/s of such invariance (cultural or non cultural) is critical. Qualitative approach might also play a great role to understand the non-invariant items better.

References

Aday, L. A., Ronald A., and Gretchen V. F., (1980). *Health Care in the U.S. Equitable for Whom?* Beverly Hills, CA: Sage.

Aiyer, A. N., Kip, K. E., Mulukutal, S. R., Marroquin, O. C., Hipps, L.Jr., Reis, S. E. (2007). Predictors of significant short-term increases in blood pressure in a community-based population, *American Journal of medicine*; 120 (960-967).

Alaimo K, Olson C.M., Frongillo E. A. (1999). Importance of cognitive testing for survey items: an example from food security questionnaires. *Journal of nutrition Education*, *31(5),* 269-275.

Anderson, R. J., Freedland, K. E., Clouse, R. E., Lustman, P. J., (2001). The prevalence of comorbid depression in adults with diabetes. *Diabetes Care 24*:1069–1078.

Anderson, R. T., Aaronson, N. K., & Wilkin, D. (1993). Critical review of the international assessments of health-related quality of life. *Quality of Life Research, 2(6),* 369-395.

Arieli, A. & Ayche, S. (1993). Psychopathological aspects of the Ethiopian immigration to Israel. *Israel Journal of Medical Science 29* ,411-418.

Baker, F. M., Velli, S. A., Freidman, J., & Wiley, C. (1995). Screening tests for depression in older Black versus White patients. *American Journal of Geriatric Psychiatry, 3(1),* 43-51.

Barnes-Holmes, Y., Barnes-Holmes, D., Smeets, P. M., & Luciano, C. (2004). The derived transfer of mood functions through equivalence relations. *The Psychological Record, 54*, 95-114.

Barrett, P. (2007). Structural equation modeling: Adjudging model fit," *Personality and Individual differences*, *42 (5),* 815-824.

Beck, A.T., Ward, C.H., Mendelson, M., et al. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561-571.

Beckstead, J. W., Yang, C. Y., & Lengacher, C. A. (2008). Assessing cross-cultural validity of scales: a methodological review and illustrative example. *International Journal of Nursing Studies*, *45*(1), 110–119.

Behling, O. & Law, K.S. (2000). *Translating Questionnaires and Other Research Instruments: Problems and Solutions*. Thousand Oaks, CA: Sage Publications.

Beiser M (1990) Migration: Opportunity or mental health risk. *Triangle.* 29. 83-90.

Bemak, F., Chung, R. C-Y., & Pedersen, P.B. (2003). *Counseling refugees: A psychosocial approach to innovative multicultural interventions*. Westport, CT: Greenwood Press.

Bentler, P., M. (1990). Comparative Fit Indexes in Structural Models, *Psychological Bulletin*, *107 (2),* 238-246.

Bentler, P., M. & Bonnet, D.,C. (1980). Significance Tests and Goodness of Fit in the Analysis of Covariance Structures, *Psychological Bulletin*, *88 (3),* 588-606.

Bhugra, D. (2003) Migration and depression. *Acta Psychiatrica Scandinavica Supplementum*, *418*, 67–73.

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

Boutin-Foster, C. (2008). An item-level analysis of the Center for Epidemiologic Studies Depression Scale (CES-D) by race and ethnicity in patients with coronary artery disease. *International Journal of Geriatrics Psychiatry, 23(10),* 1034-1039.

Bravo, M. (2003). Instrument development: Cultural adaptations for ethnic minority research. In G. Bernal, J. E. Trimble, A. K. Burlew, & F. T. L. Leong (Eds.), *Handbook of racial and ethnic minority psychology* (pp.220-236). Thousand Oaks, CA: Sage Publications.

Brislin, R. W., Lonner, W. & Thorndike, R. (1973). *Cross-cultural methods*. New York: Wiley.

Byrne, B.M. (1998), *Structural equation modeling with LISREL, PRELIS and SIMPLIS: Basic concepts,applications and programming*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology, 34*, 155-175.

Callahan, C. M., & Wolinsky, F. D. (1994). The effect of gender and race on the measurement properties of the CES-D in older adults. *Medical Care, 32*, 341-356.

Canino, G., Lewis-Fernandez, R., & Bravo, M. (1997). Methodological challenges in cross-cultural mental health research. *Transcultural Psychiatry, 34*, 163-184.

Carbone E.,T., Cambell M.,K., Honess-Morreale L (2002). Use of congnitive interview techniques in the development of nutrition surveys and interactive nutrition messages for low-income populations. *Journal of American Diet Association*, *102*, 690-696.

Centers for Disease Control and Prevention (CDC). Web-based Injury Statistics Query and

   Reporting System (WISQARS) (2005). National Center for Injury Prevention and

   Control, CDC (producer). Retrieved on March, 15/2010 from: www.cdc.gov/

   ncipc/wisqars/default.htm.

Chan, K. S., Orlando, M., Ghosh-Dastidar, B., Duan, N., & Sherbourne, C. D. (2004). The

   interview mode effect on the Center for Epidemiolgical Studies Depression (CES-D)

   scale; an item response theory analysis; *Medical Care, 42(3);* 281 - 289.

Chen, J., Ng, E., & Wilkins, R. (1996). The health of Canada's immigrants in 1994–95. *Health*

   Reports, *7*(4), 33–45.

Cheng, A. T. (2001). Case definition and culture: Are people all the same? *British Journal of*

   Psychiatry, 179, 1-3.

Cheung, C. & Bagley, C. (1998) Validating an American scale in Hong Kong: Center for

   Epidemiological Studies Depression Scale (CES-D). *Journal of Psychology*, *132,* 169–

   186.

Chiriboga, D.A., Jang, Y., Banks, S. & Kim, G. (2007) Acculturation and its effects on

   depressive symptom structure in a sample of Mexican American elders. *Hispanic Journal*

   of Behavioral Sciences, 29, 83–100.

Cho, M. J., Nam, J. J., & Suh, G. H. (1998). Prevalence of symptoms of depression in

   nationwide sample of Korean adults. *Psychiatry Research, 81*, 341-352.

Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R.

   H. Hoyle (Ed.). *Structural equation modeling* (pp. 37 – 55). Thousand Oaks, CA: Sage.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing

   measurement invariance. *Structural Equation Modeling, 9(*2), 233-255.

Chung, R. C., Bemak, F., Ortiz, D. P., & Sandoval-Perez, P. (2008). Promoting the mental health
of immigrants: A multicultural/social justice perspective. *Journal of Counseling and
Development*, *86*, 310–317.

Cole, S. R., Kawachi, I., Maller, S. J., & Berkman, L. F. (2000). Test of item-response bias in the
CES-D scale: Experience from the New Haven EPESE study. *Journal of Clinical
Epidemiology, 53*, 285-289.

Collins, J. W. Jr. & Shay, D. K. (1994). Prevalence of low birth weight among Hispanic infants
with United States-born and foreign-born mothers: The effect of urban poverty. *American
Journal of Epidemiology, 139*(2), 184-192.

Comstock. G. W. & Helsing. K. (1976). Symptoms of depression in two communities.
Psychological Medicine, *6(4)*, 551- 563.

Crockett, L. J., Randall, B. A., Shen, Y.-L., Russel, S. T., & Driscoll, A. K. (2005). Measurement
equivalence of the Center for Epidemiological Studies Depression Scale for Latino and
Anglo adolescents: A national study. *Journal of Consulting and Clinical Psychology, 73*,
47-58.

Devins, G. M., and Orme, C.M., (1986). Center for Epidemilogic Studies Depression Scale. In
Kayser, D. J., and Sweetland, R. C., (Eds). *Test critiques 2*: 144-160. Kansas City:Test
corporation of America, Inc.

Devins, G. M., Orme, C. M., Costello, C. G., Binik, Y. M., Frizzell, B., Henderikus, J. S., Pullen,
W. M. (1988). Measuring depressive symptoms in illness populations: Psychometric
properties of the Center for Epidemiologic Studies Depression (CES-D) Scale. *Journal of
Health Psychol ogy* , *2*,139–156.

Dierker, L. C., Albano, A. M., Clarke, G. N., Heimberg, R. M., Kendall, P. C., Meri- kangas, K. R., et al. (2001). Screening for anxiety and depression in early adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 929–936.

Diamantopoulos, A. & Siguaw, J. A. (2000). *Introducing LISREL*. London: Sage Publications.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation**. *Measurement and Evaluation in Counseling and Development, 43*(2), 121-149.

Edman, J. L., Danko, J. P., Andrade, N., McArdle, J. J., Foster, J.,& Glipa J. (1999). Factor structure of the CES-D (Center for Epidemiologic Studies Depression Scale) among Filipino-American adolescents, *Social Psychiatry and Psychiatric Epidemiology, 34* , 211–215.

Everitt, B. S. (1975). Multivariate analysis: The need for data, and other problems. *British Journal of Psychiatry, 126*:237–240.

Fava, G. A. (1983). Assessing depressive symptoms across cultures: Italian validation of the CES-D self-rating scale. *Journal of Clinical Psychology*, *39*, 249–251.

Fenta, H., Hyman, I., & Noh, S. (2004). Determinants of depression among Ethiopian immigrants and refugees in Toronto. *Journal of Nervous and Mental Disease, 192*, 363-372.

Flaherty, J. A., Gaviria, F. M., Pathak, D., Mitchell, T., Wintrob, R., Richman, J. A., et al. (1988). Developing instruments for cross-cultural psychiatric research. *Journal of Nervous and Mental Disorder, 176*(5), 257-263.

Foley, K. L., Reed, P. S., Mutran, E. J., & DeVellis, R. F. (2002). Measurement adequacy of the CES-D among a sample of older African-Americans. *Psychiatry Research, 109*, 61-69.

Fountoulakis K., Iacovides A., Kleanthous S., et.al. (2001). Reliability, validity and

    psychometric properties of the Greek translation of the Center for Epidemiological

    Depression (CES-D) Scale. *Biomedical Central Psychiatry 1(3).*

Fuhrer, R. & Rouillon, F. (1989) The French version of CES-D: Description and translation of

    the self-report scale. *Psychiatrie and Psychobiologic*, *4*, 163–166.

Gaim, K. (1985). *African Refugees: Reflection on the African Refugee Problem.* Trenton, N.J.:

    African World Press.

Gaim, K. (1987). *Refugees and development in Africa: The case of Eritrea.* NJ: The Red Sea

    press INC.

Gaim K. (2009), *Eritrea. A Dream Defferred*, James Currey, New York: Woodbridge and

    Rochester.

Gatz, M., & Hurwicz, M-L. (1990). Are old people more depressed? Cross-sectional data on

    Center for Epidemiological Studies Depression Scale (CES-D) factors. *Psychology and*

    Aging, 5, 284-290.

Getahun S., A. (2007). *The history of Ethiopian immigrants and refugees in America 1900-2000.*

    Patters of migration, survival, and adjustment. New York. LFB Scholarly Publishing

    LLC.

Ghubash, R., Daradkeh, T. K., Al Naseri, K. S., Al Boushi, N. B. A. & Al Daheri, A. M. (2000).

    The performance of the Center for Epidemiological Studies Depression Scale in an Arab

    female community. *International Journal of Social Psychiatry*, *46,* 241–249.

Golding, J., Aneshensel, C., & Hough, R. (1991). Responses to depression scale items among

    Mexican-Americans and non-Hispanic Whites. *Journal of Clinical Psychology, 47,* 61-

    75.

Goncalves, B. & Fagulha, T. (2004). The Portuguese version of the Center for Epidemiological Studies Depression Scale (CES-D). *European Journal of Psychological Assessment*, 20, 339–348.

Gonzalez, H. M., Haan, M. N., & Hinton, L. (2001). Acculturation and the prevalence of depression in older Mexican Americans: Baseline results of the Sacramento Area Latino Study on Aging. *Journal of the American Geriatrics Society, 49*, 948-953.

Greenberg, P. E., Kessler, R. C., Birnbaum, H. G., Leong, S. A., Lowe, S. W., Berglund, P. A., Corey-Lisle, P. K. (2003). The economic burden of depression in the United States: how did it change between 1990 and 2000? *Journal of Clinical Psychiatry, 64*, 1465-1475.

Guarnaccia, P. J., Angel, R., & Worobey, J. L. (1989). The factor structure of the CES-D in the Hispanic Health and Nutrition Examination Survey: the influences of ethnicity, gender, and language. *Social Science and Medicine, 29,* 85-94.

Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology, 46*(12), 1417-1432.

Gupta, R., Punetha, D. & Diwan, S. (2006) The revised CES-D scale for caregivers of the elderly in India. *International Journal of Aging and Human Development*, *62*, 61–78.

Hambleton, K. R., & Zenisky, L. A., (2011). Translating and adapting tests for cross-cultural assessments. In Matsumoto, D. & Van De Vijver, F. J. R., (Eds), Cross-cultural research methods in psychology (pp. 46-74). New York: Cambridge University Press.

Hautzinger, M. (1988) The CES-D Scale: A depression rating scale for research in the general population. *Diagnostica*, *34*, 167–173.

Hays, J. C., Landerman, L. R., George, L. K., Flint, E. P., Koening, H. G., Land, K. C. et al. (1998). Social correlates of the dimensions of depression in the elderly. *Journal of Gerontology: Social Sciences, 53*, 31-39.

Hepner, T. R., & Conard, B. (2005). Eritrea Abroad: Critical Reflections on the Global Diaspora. . *Special Edition of the Eritrean Studies Review, 4*(2), v-xvii.

Hermansson, A., Timpka, T., & Thyberg, M. (2002). The mental health of war-wounded refugees: An 8-year follow-up. The Journal of Nervous and Mental Disease, *190*(6), 374_380.

Hernández, D., & Charney, E. (Eds.). (1998). *From generation to generation: The health and well-being of children of immigrant families.* Washington, DC: National Academy Press.

Hollifield , M., Warner, T. D., Lian , N., et al.(2002). Measuring trauma and health status in refugees: a critical review. *JAMA, 288*:611-621.

Hu, L.T. and Bentler, P.M. (1999), "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives," Structural Equation Modeling, 6 (1), 1-55.

Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural psychology. *Journal of Cross-Cultural Psychology, 16*, 131-152.

Inoba, A., Thoits, P. A., Ueno, K., Gove, W. R., Evenson, R. J., & Sloan, M. (2005). Depression in the United States and Japan: Gender, marital status, and SES patterns. *Social Sciences and Medicine, 61,* 2280-2292.

Jang, Y., Kim, G., & Chiriboga, D. A. (2005). Acculturation and manifestation of depressive symptoms among Korean American older adults. *Aging and Mental Health, 9*, 500-507.

Janssens, M., Brett, J. M., & Smith, F. J. (1995). Confirmatory cross cultural research: Testing the viability of a corporation-wide safety policy. *Academy of Management Journal*, 38 (364-382).

Jobe J. B., Mingay D. J. (1989). Congnitive research improves questionnaires. *American Journal of Public Health*. *79*, 1053-1055.

John, M. T., Hirsch, C., Reiber, T., & Dworkin, S. (2006). Translating the research diagnostic criteria for temporomandibular disorders into German: evaluation of content and process. *Journal of Orofacial Pain, 20*(1), 43-52.

Jones, E. E. & Korchin, S. J. (1982). Minority mental health perspectives. *In* E. E. Jones and S. J. Korchin (Eds.), *Minority Mental Health* (pp. 3-36). New York: Praeger.

Jöreskog, K. & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago, IL: Scientific Software International Inc.

Kanazawa, A., White, P. M. & Hampson, S. E. (2007) Ethnic variation in depressive symptoms in a community sample in Hawaii. *Cultural Diversity and Ethnic Minority Psychology*, *13*, 35–44.

Katon, W., von Korff, M., Ciechanowski, P., Russo, J., Lin, E., Simon, G., Ludman, E.,Walker, E., Bush, T., Young, B., (2004). Behavioral and clinical factors associated with depression among individuals with diabetes. *Diabetes Care 27*:914 –920.

Kim, G., Chiriboga, D. A., Jang, Y. (2009). Cultural equivalence in depressive symptoms in older White, Black, and Mexican-American adults. *Journal of the American Geriatrics Society, 57,* 790-796.

Kleinman, A.(1982). Depression and Neurasthenia in the people's Republic of China. *Culture, medicine, and Psychiatry*, 6, 1-80.

Kline, R. B. (2010), *Principles and practice of structural equation modeling* (3rd ed.). New York: The Guilford Press.

Krause, N., & Liang, J. (1992). Cross-cultural variations in depressive symptoms in later life. *International Psychiatrics, 4*(Suppl. 2), 185-202.

Krause, N., & Liang, J. (1993). Stress, social support, and psychological distress among the Chinese elderly. *Journal of Gerontology: Psychological Sciences, 48*, 282-291.

Liang, J. (2002). Assessing cross-cultural comparability in mental health among older adults. In J. H. Skinner, J. A. Teresi, D. Holmes, S. M. Stahl, & A. L. Stewart(Eds.), *Multicultural measurement in older populations* (pp.11-21). New York: Springer Publishing Company.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51*, 201-226.

MacCallum, R.C., Browne, M.W., and Sugawara, H., M. (1996), "Power Analysis and Determination of Sample Size for Covariance Structure Modeling," Psychological Methods, 1 (2), 130-49.

MacKinnon, A., McCallum, J., Abdrews, G., & Anderson, I. (1998). The Center for Epidemiologic Studies Depression Scale in older community samples in Indonesia, North Korea, Myanmar, SriLanka, and Thailand. Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, *53*, 343–352.

Marsella, A. J., Shizuru, L. S., Brennan, J. M., & V. Kameoka (1981). Depression and body image satisfaction. *Journal of Cross Cultural Psychology,* 12, 360-371.

McDonald, R. P. and Ho, M-H. R. (2002). Principles and Practice in Reporting Statistical Equation Analyses, *Psychological Methods*, *7 (1)*, 64-82.

McIntosh, C. (2006). Rethinking fit assessment in structural equation modeling: A commentary and elaboration on Barrett (2007), *Personality and Individual Differences*, 42 (5), 859-67.

Miles, J. and Shevlin, M. (1998). A time and a place for incremental fit indices. *Personality and Individual Differences*,*42* (5), 869-74.

Mills T, Henretta J.(2001). Racial, ethnic, and sociodemographic differences in the level of psychosocial distress among older Americans. *Research on Aging* ;*23*(2):131–152.

Morris, M. W., Leung, K., Ames, D., & Lickel, B. (1999). Views from inside and outside: Integrating emic and etic insights about culture and justice judgment. *Academy of Management Review, 24*, 781-796.

Murphy, J. M. (2002). Symptom scales and diagnostic schedules in adult psychiatry. In M. T. Tsuang & M. Tohen (Eds.), *Textbook in psychiatric epidemiology* (pp. 273-332). New York, NY: Wiley Liss.

Murray, C. J. L., and Lopez, A.D. (1997). Alternative projections of mortality and disability by cause 1990 – 2020: Global Burden of Disease Study. The Lancet, 349, 1498 – 1504.

Neighbors, H. W., Trierweiler, S. J., Ford, B. C., & Muroff, J. R. (2003). Racial differences in DSM diagnosis using a semi-structured instrument: the importance of clinical judgment in the diagnosis of African Americans. *Journal of Health and Social Behavior, 44*(3), 237-256.

Nguyen, H. T., Kitner-Triolo, M., Evans, M. K., & Zonderman, A. B. (2004). Factorial invariance of the CES-D in low socioeconomic status African Americans compared with a nationally representative sample. *Psychiatry Research, 126*, 177-187.

Noh, S., Avison, W.R. & Kaspar, V. (1992) Depressive symptoms among Korean immigrants: Assessment of a translation of the Center for Epidemiological Studies Depression Scale. *Psychological Assessment*, *4*, 84–91.

Noh, S., Kaspar, V., & Chen, X. (1998). Measuring depression in Korean immigrants: Assessing validity of the translated Korean version of the CES-D Scale. *Cross-Cultural Research*, *32*, 358-377.

Papassotiropoulos, A., & Heun, R. (1999). Screening for depression in the elderly: A study on misclassification by screening instruments and improvement of scale performance. *Progress in Neuro-Psychopharmacolgy and Biological Psychiatry, 23*, 431-446.

Perkonigg, A., Kessler, R. C., Storz, S., & Wittchen, H. U. (2000) Traumatic events and post-traumatic stress disorder in the community: prevalence, risk factors and co morbidity. *Acta Psychiatrica Scandinavica; 101:* 46-59.

Perreira, K. M., Deeb-Sossa, N., Harris, K. M., & Bollen, K. (2005). What are we measuring? An evaluation of the CES-D across race/ethnicity and immigrant generation. *Social Forces, 83*, 1567-1602.

Persson, G. (Ed.). (2000). Different conditions–different health. A study of immigrants from Chile, Poland, Iran and Turkey. Stockholm: The National Board of Health and Welfare. SoS report 2000:3.

Pignone M. P., Gaynes B. N., Rushton, J. L., et. al. (2002). Screening for depression in adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Annals of Intern Med*icine.*136*(10),765-776.

Poortinga, Y. H. & Van de Vijver, F. J. R. (1987). Equivalence of cross cultural data: An overview of basic issues. *International Journal of Psychology*, 24, 737-756.

Posner, S. F., Stewart, A. L., Marín, G., & Pérez-Stable, E. J. (2001). Factor variability of the

    Center for Epidemiological Studies Depression Scale (CES-D) among urban Latinos.

    Ethnicity and Health, 6, 137-144.

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.

Rait, G., & Burns, A. (1998). Screening for depression and cognitive impairment in older people

    from ethnic minorities. *Age and Ageing, 27*, 271-275.

Raskin A, Schulterbrandt J, Reating N, et al. (1967). Factors of psychopathology in interview,

    ward behavior, and self-report ratings of hospitalized depressives. *Journal of Consulting

    Psychology 31*, 270-278.

Rhee, S. H., Petroski, G. F., Parker, J. C., Marr, K. L., Wright, G. E., Multon, K. D., Buchholz,

    J. L. & Komatireddy, G. R., (1999). A confirmatory factor analysis of the Center for

    Epidemiology Studies Depression scale in rheumatoid arthritis patients: additional

    evidence for a four-factor model, *Arthritis Care Research 12* (1999), 392–400.

Riordan, C. M., & Vandenberg, R. J. (1994). A central question in cross cultural research: Do

    employees of different cultures interpret work related measures in an equivalent manner?

    Journal of Management, 20(3), 643-671.

Roberts, R. E. (1980). Reliability of the CES-D scale in different ethnic contexts. *Psychiatry

    Research, 2*, 125-134.

Roberts, R. E., Vernon, S. W., & Rhoades, H. M. (1989). Effects of language and ethnic status

    on reliability and validity of the Center for Epidemiologic Studies-Depression Scale with

    psychiatric patients. *Journal of Nervous and Mental Disease, 177*, 581-592.

Roberts, R. E., Rhoades, H. M., & Vernon, S. W. (1990). Using the CES-D scale to screen for depression and anxiety: Effects of language and ethnic status. *Psychiatry Research, 31,* 69-83.

Schaffer, B. S., & Riordan, C. M. (2003). A review of cross-cultural methodologies for organizational research: A best-practice approach. *Organizational Research Methods, 6,* 169-215.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and limitations. *Human Resources Management Review*, 18, 210-222.

Selten, J. P., Sijben, N. (1994). First admission rates for schizophrenia in immigrants to the Netherlands. The Dutch National Register. *Social Psychiatry and Psychiatric Epidemiology*,*29*, 71–77

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.

Shalev, A. Y., Freedman, S., Peri, T., Brandes, D., Sahar, & T., Orr, S. P., et al. (1998). Prospective study of posttraumatic stress disorder and depression following trauma. *American Journal of Psychiatry; 155*: 630-637.

Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W.R. (2005), "A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models," *Journal of Business Research*, *58*(1), 935-43.

Shima S, Shikano T, Kitamura T., & Asai M. (1985). New self-rating scale for depression. *Clinical Psychiatry, 27* (717-723).

Steenkamp, J. B.E.M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78-90.

Steiger, J. H., & Lind, J.C. (1980). Statistically-based tests for the number of common factors. Paper presented at the annual Spring Meeting of the Psychometric Society in Iowa City. May 30, 1980.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling, *Personality and Individual Differences*, *42*(5), 893-98.

Sundquist J., Behmen-Vincevic, A., Johansson, S. E. (1898). Poor quality of life and health in young to middle aged Bosnian female war refugees: A population based study. *Public Health, 112*, 21–26.

Swenson, C. J., Baxter, J., Shetterly, S. M., Scarbro, S. L., & Hamman, R. F. (2000). Depressive symptoms in Hispanic and Non-Hispanic White rural elderly: the San Luis Valley Health and Aging Study. *American Journal of Epidemiology, 152*, 1048-1054.

Tabachnick, B. G., & Fidell, L. S. (2007), *Using multivariate statistics* (5th ed.). New York: Allyn and Bacon.

UNCHR (2009). World Report Eritrea. Retrieved on 03/10/2010 from: http://www.unhcr.org/refworld/country,,,,ERI,,49705fa350,0.html.

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*, 4-70.

Van de Vijver, F. (2001). The evolution of cross-cultural research methods. In D. Matsumoto (Ed.), *The handbook of culture and psychology* (pp.77-97). New York: Oxford University Press.

Van Tran, T. (1997). Exploring the equivalence of factor structure in a measure of depression between Black and White women: Measurement issues in comparative research. *Research on Social Work Practice, 7*, 500-517.

Vazquez, F. L., Blanco, V., & Lopez, M. (2007). An adaptation of the Center for Epidemiologic Studies Depression Scale for use in non-psychiatric Spanish populations. *Psychiatry Research, 149*(1-3), 247-252.

Vega, W. A., Kolody, B., Aguilar-Gaxiola, S., Alderete, E., Catalano, R., Caraveo-Anduaga , J. (1998) Lifetime prevalence of DSM-III-R psychiatric disorders among urban and rural Mexican Americans in California. *Arch Gen Psychiatry, 55*:771-778.

Wiesinger, G. F., Nuhr, M., Quittan, M., Ebenbichler, G., Wolfl, G., & Fialka-Moser, V. (1999). Cross-cultural adaptation of the Roland-Morris questionnaire for German-speaking patients with low back pain. *Spine (Phila Pa 1976), 24*(11), 1099-1103.

Woldemichael. (1998). Eritrean and Ethiopian Refugees in the United States. *Eritrean Studies Review, 2*(2), 89-109.

Weissman M. M., Sholomaskas D., Pottenger M., Prusoff, B., A., & Locke, B., Z. (1977). Assessing depressive symptoms in five psychiatric populations: a validation study. *American Journal of Epidemiology, 106*(3), 203-214.

WHO (nd). World health organization: Depression. Retrieved from: http://www.who.int/mental_health/management/depression/definition/en/ on 03/04/2010.

WHO. (2008). Depression. Retrieved May 16,2011, from http://www.who.int/mental_health/management/ depression/definition/en/.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of Book psychological instruments: Applications in the substance use domain. In K. J. Chapter

Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281-324). Washington, DC: American Psychological Association.

Wren, K. and Boyle P. (2001). Migration and Work-Related Health in Europe: A Literature Review, SALTSA Report 2001:1, Stockholm.

Yu, D.S.F., Lee, D.T.E, & Woo, J. (2004). Issues and challenges of instrument translation. *Western Journal of Nursing Research, 26* (3), 307-320.

Zung W. W. K. (1965). A self-rating depression scale. *Archives of General Psychiatry, 12* (63-70).

Appendices

Appendix A: Tigrigna Version CES-D Scale

ኣብ ታሕቲ ተዘርዚሮም ዘለዉ ሙሉኣት ሓሳባት፡ ካብ ዝሓለፈ ሰሙን ክሳብ ሎሚ መዓልቲ ኣብ ዘሎ ጊዜ፣ ንዝተሰማዓካ/ኪ ወይ ዘንጸባረቕካዮ/ኪዮ ባህርያት ክሳብ ኪኖዶይ ከምዝገልጽዎ ንምሕባር፣ ብ የማን ካብ ዝርከባ ሰናዱቕ ሓንቲ እንዳመረጽካ/ኪ ሓንጽጽ/ጺ።

		ብፍጹም (ትሕቲ 1 መዓልቲ)	ኣዝዩ ውሑድ ግዜ (1-2 መዓልቲ)	ብመጠኑ (3-4 መዓልቲ)	መብዛሕታኡ ግዜ (5-7 መዓልቲ)
1.	ኣሻቐሎምኒ ዘይፈልጡ ነግራት የሻቕሉኒ ነይሮም	☐	☐	☐	☐
2.	ሸውሃተይ ተዓጽዩ፡ ናይ መግቢ ድሊይተይ ዲኩም ነይሩ	☐	☐	☐	☐
3.	ወላ ብሓገዝ በተሰብ ይኹን መቕርብ ካብ ጭንቀት ኪናገፍ ኣይከኣልኩን	☐	☐	☐	☐
4.	ማዕረ ሰበይ እየ ዝብል ስምዒት የሕዲረ	☐	☐	☐	☐
5.	ትኩር ኮይነ ክሰርሕ ኣይከኣልኩን	☐	☐	☐	☐
6.	ቃዚኔ	☐	☐	☐	☐
7.	ነብስ ወከፍ ንጥፈት ዓቢ ብድሆ ኮይኑኒ	☐	☐	☐	☐
8.	ብዛዕባ መጻእየይ ትስፉው ነይረ	☐	☐	☐	☐
9.	ሕሉፍ ሂወተይ ዘይዕዉት ኮይኑ ተረኣዩኒ	☐	☐	☐	☐
10.	ፍርሒ ተሰሚዑኒ	☐	☐	☐	☐
11.	ድቃሰይ ምዕልባጥ ዝበዝሖ ነይሩ	☐	☐	☐	☐
12.	ሕጉስ ነይረ	☐	☐	☐	☐
13.	ከም ወትሩ የዕልል ኣይነበርኩን	☐	☐	☐	☐
14.	ጸምዎኔ፡ ጸገዒ ኣልቦ ዝኮንኩ ኮይኑ ተሰሚዑኒ	☐	☐	☐	☐
15.	ሰባት ሕያዎነት ኣይለገሱለይን	☐	☐	☐	☐
16.	ጽቡቕ ህይወት ኣሕሊፈ	☐	☐	☐	☐
17.	ጽቡቕ ህይወት ኣሕሊፈ	☐	☐	☐	☐
18.	ሓዘን ተሰሚዑኒ	☐	☐	☐	☐
19.	ሰባት ዝጸልኡኒ ኮይኑ ተሰሚዑኒ	☐	☐	☐	☐
20.	ሰልቸዩኒ	☐	☐	☐	☐

Appendix B: English Version CES-D Scale

ID#: _____   DATE:_____/_____/_____

CES-D

For each statement below, check the box which best describes how often you felt or behaved this way--**DURING THE PAST WEEK, INCLUDING TODAY.**

DURING THE PAST WEEK:	None of the Time	A Little of the Time	A Moderate Amount of Time	Most of the Time
1. I was bothered by things that usually don't bother me	0	1	2	3
2. I did not feel like eating; my appetite was poor	0	1	2	3
3. I felt that I could not shake off the blues even with help from my family or friends	0	1	2	3
4. I felt that I was just as good as other people	0	1	2	3
5. I had trouble keeping my mind on what I was doing	0	1	2	3
6. I felt depressed	0	1	2	3
7. I felt that everything I did was an effort	0	1	2	3
8. I felt hopeful about the future	0	1	2	3
9. I thought my life had been a failure	0	1	2	3
10. I felt fearful	0	1	2	3
11. My sleep was restless	0	1	2	3
12. I was happy	0	1	2	3
13. I talked less than usual	0	1	2	3
14. I felt lonely	0	1	2	3
15. People were unfriendly	0	1	2	3
16. I enjoyed life	0	1	2	3
17. I had crying spells	0	1	2	3
18. I felt sad	0	1	2	3
19. I felt that people disliked me	0	1	2	3
20. I could not "get going"	0	1	2	3

PT 12/30/2003