EVALUATING THE 8-ITEM PATIENT HEALTH QUESTIONNAIRE’S PSYCHOMETRIC PROPERTIES AMONG MEN AND WOMEN OF MEXICAN DESCENT

A thesis submitted in partial fulfillment of the requirements For the degree of Master of Arts in Psychology General Experimental

by

David Alpizar

May 2015
The thesis of David Alpizar is approved:

Andrew T. Ainsworth, Ph.D.  
Date

Scott W. Plunkett, Ph.D.  
Date

Luciana Lagana, Ph.D., Chair  
Date

California State University, Northridge
DEDICATION

This thesis is dedicated to my mom, my sister, Parthenia, Dr. Lagana, Dr. Zavala, Dr. Plunkett, and Dr. Ainsworth. Each one of these persons has played a critical role in my development as a scientist and in my academic growth. I am really blessed that I came across them in my path because I was able to acquire the adequate emotional support, research skills, and knowledge in order to succeed in CSUN’s General Experimental Psychology Master’s program. Plus, the skills I have acquired will benefit me in order to conduct high quality research in the future.
ACKNOWLEDGMENT

I would like to thank my committee members who supported my efforts in writing this thesis. Each one of them has provided support in APA format and statistical mentorship, and conceptual contribution to my thesis. With their support, writing my graduate thesis was a fun and amazing experience.

To my chair, Dr. Luciana Lagana, 
To Dr. Scott Plunkett, 
To Dr. Andrew Ainsworth,  
To Dr. Sarah Berzenski,  
To Marielena Zavala,  
To Parthenia Hosh,  
To Goli Alamdari.
TABLE OF CONTENTS

Signature Page ii
Dedication iii
Acknowledgment iv
Abstract vi

CHAPTER I – INTRODUCTION 1
Statement of the Problem 5
Hypotheses 6
Assumptions 7

CHAPTER II – REVIEW OF THE LITERATURE 8
Depression 8
Theoretically Related Variables 10
Depression Measures 11
Patient Health Questionnaire 13
Development 13
Different Versions 14
Psychometric Properties 15

CHAPTER III – METHODOLOGY 21
Procedure and Sample 21
Measures 21
Statistical Analyses 23
Hypothesis 1 23
Hypothesis 2 24
Hypothesis 3 25

CHAPTER IV – RESULTS 27
Confirmatory Factor Analysis of the PHQ-8 27
Concurrent Validity of the PHQ-8 29
Convergent Validity of the PHQ-8 30

CHAPTER V – DISCUSSION 31
Summary of Findings and Discussion 31
Limitations and Research Implications 33
Implication and Conclusion 34

REFERENCES 35

APPENDIX – TABLES 47
ABSTRACT

EVALUATING THE 8-ITEM PATIENT HEALTH QUESTIONNAIRE’S PSYCHOMETRIC PROPERTIES AMONG MEN AND WOMEN OF MEXICAN DESCENT

by

David Alpizar

Master of Arts in Psychology
General Experimental

The current study examined the factor structure, concurrent, and convergent validity of the 8-item Patient Health Questionnaire (PHQ-8) among emerging adult men and women of Mexican descent residing in the United States (i.e., U.S. born Mexican Americans and not U.S. born Mexicans). The participants ($M = 19.2$, $SD = 1.6$, 18-29 years of age) were recruited from the psychology subject pool using self-reported online or paper-pencil surveys. A confirmatory factor analysis compared one factor, two factor, and bi-factor models. The one factor structure was the best fit among men, whereas the bi-factor model structure was the best fit for women. Nevertheless, both of these findings indicated that the PHQ-8 is unidimensional, given that items are driven by the general factor in the one factor and bi-factor model. Depression (as measured in the PHQ-8) was significantly correlated with the previously validated CES-D short version and theoretically related variables (i.e., quality of life, parental support, and perceived stress). These findings indicate that the PHQ-8 has excellent concurrent and convergent validity. Implications of the findings are discussed.
CHAPTER I
INTRODUCTION

Depression is characterized by depressed/irritable mood, absence of interest or pleasure, change in appetite and sleep routines, fatigue, feelings of guilt/worthlessness, problems concentrating, and suicidal thoughts (American Psychiatric Association [APA], 2013; World Health Organization, 2007). Depression is a disorder that can impact a person’s wellbeing and daily functioning (Oquendo, Lizardi, Greenwald, Weissman, & Mann, 2004) and worsen current medical conditions such as cardio-vascular disease (American Heart Association, 2013; Richardson et al., 2003; World Health Organization, 2007). Also, depression has comorbidity with other mental disorders (e.g., substance abuse disorders, Lehmann, Hubbard, & Martin, 2001), and it can potentially lead affected individuals to consider suicide (Oquendo et al., 2004). Hence, depression decreases the overall wellbeing of the affected individuals.

Emerging adults (18 to 29 years of age, Arnett, 2011), including college students (Eisenberg, Gollust, Golberstein, & Hefner, 2007; Walsemann, Gee, & Geronimus, 2009), are at risk for developing depression because they face challenges in their personal development (e.g., finding a job) and social development (e.g., finding romantic partners) (Adkins, Wang, Dupre, Oord, & Elder, 2009; Arnett, 2011; Hill, Jackson, Roberts, Lapsley, & Brandenberger, 2011). Among emerging adults, there are ethnic differences on the prevalence rate of depression. For instance, individuals of Mexican descent have reported higher levels of depression than European Americans and other non Spanish speaking ethnic groups (Centers for Disease Control and Prevention [CDC], 2010; Roberts, Roberts, & Chen, 1997). Regardless of these differences, there is a demand for
more reliable and valid depression inventories for individuals of Mexican descent because the literature lacks research on this topic.

The Patient Health Questionnaire (PHQ-9) is a relatively new instrument (Spitzer et al., 1999). This tool contains 9 items quantifying depression. The response choices range from 0 (not at all) to 3 (nearly every day). There are two features to highlight. First, the PHQ-9 is short and quick to score (Spitzer et al., 1999). Second, the items parallel the depression diagnostic criteria from the DSM-V (APA, 2013). Table 1 (see Appendix) illustrates this similarity.

Since the PHQ-9 has the above-mentioned characteristics, it has become widely used over the last few years for screening and research for depression (Kroenke, Spitzer, Williams, & Löwe, 2010). Various short forms of the PHQ-9 have been developed, such as the PHQ-8 (8 items) and the PHQ-2 (2 items). The PHQ-8 and PHQ-9 are exactly the same, except the PHQ-8 does not include the 9th item. The current study uses the PHQ-8 over the PHQ-9 for the following reasons. First, the 9th item is a doubled-barreled item. The 9th item’s context is “Thoughts that you would be better off dead or hurting yourself in some way.” The context of the item might create confusion among the clients.
since it is measuring two concepts simultaneously (Crede’, 2010). Since the 9th item lacks clarity, the responses to it do not indicate whether a person is more suicidal or more hurting his/herself. For this reason, researchers have recommended that double-barreled items must be treated with cautious (Crede’, 2010).

Secondly, the PHQ-8 and PHQ-9 are highly related (Kroenke et al., 2010; Razykov, Ziegelstein, Whooley, & Thombs, 2012). Studies have documented the PHQ-8 and the PHQ-9 have identical scoring thresholds for depression (Kroenke & Spitzer, 2002; Kroenke, Strine, Spitzer, Williams, Berry, & Mokdad, 2009), indicating similar sensitivity and specificity. This suggests that the PHQ-8 and PHQ-9 measure the same depression latent variable using similar scoring criteria, thus it is not a surprise that researchers and clinicians are using the PHQ-8 as well. Because the 9th item lacks clarity and removing it does not jeopardize the PHQ-8’s psychometric properties, in the current study the 9th item was not included.

Although the PHQ-8 might have very prosperous usage in a research and clinical setting, the scale was initially developed ($n = 3000$) and evaluated on samples compromised mostly of European Americans (Spitzer et al., 1999). This lack of cultural sensitivity needs to be remediated, as was attempted in the present study. Having unreliable and invalid scales might lead to inaccurate self-report and not capture the target construct, so it is critical to examine its psychometric properties (Banks, 1995; Schwarz, 1999). For this concern, the PHQ-8 must be evaluated among ethnic groups.

Additionally, there are mixed published findings about the factor structure of the PHQ-8. On one end, researchers found that the PHQ-8’s items were driven by one factor among Latinos (Huang, Chung, Kroenke, Delucchi, & Spitzer, 2006; Merz, Malcarne,
Roesch, Riley, & Sadler, 2011) and men of Mexican descent (Donlan & Lee, 2010). On the other end, the unidimensionality of the PHQ-8 is not fully established, given that studies report that its items are driven by two factors (i.e., affective and somatic) among the same population. For example, Granillo (2012) found that two factors was good fit for this scale once two of the items were dropped. Due to these inconsistencies, research should be conducted to evaluate the PHQ-8’s factor structure and invariance across ethnic groups such as Latinos/as (Vega & Rumbaut, 1991). Otherwise, there is the probability that self-reported responses are underestimated (Chen, 2008).

In addition to establishing factor structure and invariance, it is critical to examine the PHQ-8’s validity. Concurrent validity is the relationship between a tested inventory and a previous validated inventory documented to measure the same construct in the literature (Gay, 1996). In this case, the PHQ-8 must be examined with a well-established depression inventory in order to show it is capturing the same construct. Second, it is important to evaluate the PHQ-8’s convergent validity. Convergent validity is the degree to which an inventory relates to theoretically related variables (Gay, 1996). The following variables have been documented to relate to depression in the literature: parental support, perceived stress, and quality of life. Parental support refers to the amount of involvement and encouragement that parents provide to their offspring (Nebitt & Lombe, 2008). Scholars have widely documented the important role that parental support has on depression, as individuals who receive more parental support have lower levels of depression (e.g., Liem, Lustig, & Dillon, 2010; Needham, 2008; Yeung & Leadbeater, 2010). Additionally, perceived stress is defined as an unexpected situation/social challenge (e.g., car accident) where negative feelings/behaviors (e.g., anger, irritation) are
produced (Ng & Jeffery, 2003). Multiple sources have found that individuals who experienced stressful situations have higher symptoms of depression (Flores, Tschann, Dimas, Bachen, Pasch, & de Groat, 2008; Ng & Jeffery, 2003; Vega, Rodriguez, & Ang, 2010. Quality of life corresponds to the overall wellbeing (e.g., mental, physical, social, and emotional health) of an individual (Husson, Mols, & Van de Poll-Franse, 2010). Several studies have shown that higher levels of depression negatively impact wellbeing (Abraham & Fava, 1999; Gotlib, 1997). When individuals are depressed, they find less pleasure in activities, experience more physical fatigue, and have lower self-worth, all of which can undermine quality of life. Because the importance of parental support, perceived stress, and quality of life are widely documented in the literature, the relationship of these theoretically related variables with depression must be examined to evaluate convergent validity.

**Statement of the Problem**

Depression can be detrimental to an individual’s well being (Abraham & Fava, 1999; Gotlib, 1997; Oquendo et al., 2004). As already mentioned, the 8-item Patient Health Questionnaire (PHQ-8) is a self-administered instrument used to assess mental disorders; it is widely used in research and clinical practice to measure depression (Kroenke et al., 2010). However, very few researchers have examined its psychometrics properties on emerging adult men and women of Mexican descent. To be more specific, there was only one study that targeted emerging adults of Mexican descent, who were mainly grouped with other Spanish speaking groups as Latinos/as or Hispanics. There is available evidence that different Latino groups report differing levels of depression (e.g., Oquendo et al., 2004), thus there are intragroup differences among these Spanish
speaking groups. It is crucial to separate these groups when studying depression. In addition, depression inventories’ psychometric properties should be examined across these Spanish speaking groups. If unreliable and invalid scales are used to screen depression within ethnic groups, they could not capture the desired construct and could lead to reporting inaccurate results.

**Purpose**

In the current study, I examined the PHQ-8’s psychometrical properties among emerging adults of Mexican descent. To be more specific, I aimed to examine the PHQ-8’s factor structure, factor invariance, concurrent validity, and convergent validity across the target populations.

**Hypotheses**

**Research Hypotheses**

Based on the review of literature included in Chapter 2, the following research hypotheses were developed.

1. Hypothesis 1 aimed to establish the factor structure of the PHQ-8. The literature has provided mixed findings about the factor structure (unidimensional vs. multidimensional) of this inventory; therefore I examined and established this factor structure among emerging adult men and women of Mexican descent.

   1a. It was expected that the PHQ-8 would be driven by one latent variable across these groups.

2. Hypothesis 2 aimed to test the factor invariance of the PHQ-8 between men and women of Mexican descent.

   2a. A multi-group confirmatory factor analyses would be conducted to establish
configural invariance, weak factorial invariance, strict factorial invariance, latent variable’s variance, and equal factor means across the groups (i.e., men vs. women); thus, full factor invariance would be tested.

3. Hypothesis 3 aimed to examine the validity of the PHQ-8 among men and women of Mexican descent.

3A. Concurrent validity would be assessed by calculating the correlation of the PHQ-8 with the Center for Epidemiological Scale of Depression (CES-D).

3B. Convergent validity would be tested by establishing the relationship between the PHQ-8, stress, parental support, and quality of life. Strong relationships are expected between the variables, thus establishing strong convergent validity of the PHQ-8.

Assumptions

This research study was created based upon certain assumptions. It was assumed that participants were fluent English Speakers because they lived in the United States and were able to comprehend the consent form. Moreover, it was assumed that participants self-reported data honestly and accurately, and that selected inventories were adequate for the participants’ ethnic group.
CHAPTER II
REVIEW OF LITERATURE

Depression

According to the World Health Organization (2012), depression is characterized by depressed/irritable mood, absence of interest or pleasure, change in appetite and sleep routines, fatigue, and feelings of guilt/worthlessness, problems concentrating, and suicidal thoughts. In the Diagnostic Statistical Manuals of Mental Disorders (DSM) 5 (APA, 2013), major depression is characterized by the symptoms described upon, yet it impairs daily function. Also, symptoms are current in the last two weeks.

It is predicted that, by 2030, depression will be the second leading cause of disease and disability in the world (Lopez & Mathers, 2006). Currently, it is the third leading cause of disease and disability, especially among women (CDC, 2010). Many studies have documented that Latinos (including Mexican Americans) have higher levels of depression and sadness than their non-Latino peers (Granillo, 2012; Knopf, Park, & Mulye, 2008; Oquendo et al., 2004; Paxton, Valois, Watkins, Huebner, & Drane, 2007). To be more specific, individuals of Mexican descent have reported higher levels of depression than European Americans (CDC, 2010) and other ethnic groups (e.g., African Americans, Roberts et al., 1997). Latino women (e.g., women of Mexican descent) have higher rates of depressive symptomatology than other ethnic groups (Brown, Meadows, & Elder, 2007; Roberts et al., 1997). Regardless of these differences, there must be more reliable and valid screening tools for measuring depression across these ethnic groups.

Accurately screening for depression is vital because it is related to daily functioning and health problems. Depression can impair the overall quality of life by
impairing daily regular functions (e.g., doing homework or going to work) to such extremes that affected individuals could potentially consider suicide (Oquendo et al., 2004). According to the World Health Organization (2012), approximately 1 million people die due to suicide yearly, which was attributed to depression and other health problems. In addition, there is comorbidity between depression and other mental disorders. For instance, Lehmann et al. (2001) reported that there is a relationship between depression and substance abuse disorders. In addition, depression plays a role in physical health related problems. Moreover, depressed people who have physical chronic diseases (e.g., asthma, angina, arthritis, diabetes) also typically have poorer levels of health than those who have physical chronic diseases than those who only have depression (World Health Organization, 2007). Furthermore, depression is related to obesity (Richardson et al., 2003), therefore, individuals who are obese and depressed tend to be more likely to develop heart diseases and other health issues (American Heart Association, 2013).

There are certain age groups that are at high risk of developing depression. Emerging adults (18 to 29 years of age) are vulnerable to developing depression, including college students (Eisenberg et al., 2007; Walsemann et al., 2009) because they typically face challenges in their personal development (e.g., finding a job, identity) and social (e.g., finding romantic partners; Adkins et al., 2009; Arnett, 2011; Hill et al., 2011). These depressed emerging adults often avoid healthy activities (e.g., social activities) and personal care (e.g., substance abuse; Oquendo et al., 2004). Thus, depression adds more burden to individuals’ already challenging situations and development, consequently decreasing their overall well-being. For these reasons, there is
a demand and an ethical obligation to have reliable and valid psychometrical inventories for depression to be used in culturally sensitive manners with individuals from all types of ethnic backgrounds.

**Theoretically Related Variables**

Quality of life is the overall well-being (i.e., mental, physical, social, and emotional health) of the individual (Husson et al., 2010). Emerging adults face great changes in their personal and social development in their transition to adulthood (e.g., finding romantic partners), and are therefore at increased risk of experiencing psychological problems such as depression (Adkin et al., 2009; Arnett, 2011; Hill et al., 2011). Depression also negatively impacts an individual’s daily ability to process information (Gotlib, 1997) and performance (Abraham & Fava, 1999); hence, it is detrimental to the overall quality of life of emerging adults.

Parental support and perceived stress play a role in the development and maintenance of depression. Parental support is the amount of involvement and encouragement that parents provide to their children (Nebitt & Lombe, 2008). The literature has widely documented the important role that this variable has on wellbeing and depression. Individuals who receive more parental support have lower levels of depression (Liem et al., 2010; Needham, 2008; Yeung & Leadbeater, 2010). Thus, parental support can diminish the risk of depression.

Perceived stress is defined as an unexpected situation/social challenge (e.g., car accident) where negative feelings/behaviors (e.g., anger, irritation) are produced (Ng & Jeffery, 2003). Multiple sources have found that individuals who experienced stressful situations have higher symptoms of depression, declined cognitive processes, and
increased maladaptive behaviors (Flores et al., 2008; Ng & Jeffery, 2003; Vega et al., 2010). Thus, perceived stress can stimulate mild-to-severe depressive symptoms.

**Depression Measures**

Due to the impact that depression often has in worsening physical and mental wellbeing, the U.S. Preventive Services Task Force (2002) recommends that adults who are patients in primary care settings get screened for depression. Screening is not diagnosing major depression, but rather it quantifies the severity of an individual’s depression (Pignone et al., 2002). Specifically, screening identifies individuals who are reporting mild to severe depressive symptoms. Once these individuals are identified, they can be referred for a psychological evaluation and, if necessary, get treatment (Vahle, Andresen, & Hagglund, 2000). Screening requires standardized scales. These measurements must be reliable and valid. For example, Chen (2008) examined the outcomes of having unreliable and invalid measurements. Chen found that measurements are not accurate when they do not capture the construct they are intended to. Hence, these unreliable and invalid scales are not useful for research or clinical testing.

One major concern is that many of the currently available depression inventories have been developed and validated on mainly European American samples (Allen & Walsh, 2000). This might lead to inaccurate self-report, given that the item’s context has the potential to have different meaning across ethnic groups (Schwarz, 1999). For example, ‘I have crying spells or feel like it’ is an item from the Zung Self-Rating Depression Scale (Zung, 1965). The expression ‘crying spells or feel like it’ means extreme sadness for one ethnic group. However and importantly, some ethnic groups might not associate this expression with sadness, or this expression is unknown to them.
This is a critical culturally-relevant psychometric issue because having measurements that are not validated across ethnic groups can result in inconsistent self-reported data (Banks, 2003; Banks, 1995). Thus, some depression inventories might not be capturing the same construct across ethnic groups. Therefore, the psychometric properties of depression measurements, specifically of the PHQ-8, need to be evaluated with different ethnic groups (Banks, 1995).

Regarding the above-mentioned concern, there are some standardized scales measuring depression that have shown good psychometric properties across diverse ethnic groups. The Center for Epidemiological Scale Depression (CES-D), the Beck Depression inventory, and the Patient Health Questionnaire are three of the most widely used scales in research and clinical testing (Fuentes & Aranda, 2012; Gloria, Castellanos, Kanagui-Munoz, & Rico, 2012; Vahle et al., 2000). The Patient Health Questionnaire will be discussed in more detail in the next section.

The Beck Depression Inventory (BDI) was designed in 1961 and contains 21 items (Vahle et al., 2000). The response choices of the BDI range from 0 (not severe) to 3 (most severe). The total score ranges from 0 to 63. Higher scores indicate the presence of more depressive symptoms. The BDI is widely used, and has been translated into other languages, such as Korean, Spanish, and French (Boyd, Le, & Somberg, 2005). In addition, it has had a couple of revisions such as the BDI-II and BDI-Short Form. The BDI and BDI-II have shown good reliability and validity, and factor invariance was met with diverse ethnic groups and age groups (Contreras, 2004; Joe, Woolley, Brown, Ghahramanlou-Holloway, & Beck, 2008).

The Center for Epidemiological Scale Depression (CESD) has 20 items.
Participants select from four possible responses with response choices ranging from 0 (rarely or none of the time) to 4 (most of all the time). The total score ranges from 0 to 60 where higher scores show the presence of more depressive symptoms. It has shown good reliability and validity. Additionally, its factor invariance has been supported across many groups, such as diverse ethnic groups, age groups, and social economic status (Crockett, Randall, Shen, Russell, & Driscoll, 2005; Nguyen, Kitner-Triolo, Evans, & Zonderman, 2004; Radloff, 1977). Its factor structure contains depressed affect, positive affect, somatic activity, and an interpersonal factor. Like the BDI, it has been translated into many languages and has shorter versions (i.e., the CESD Short Form). It is extensively used in research and clinical settings (Crockett et al., 2005). The 10-item version of this scale was used in the present study. This CESD Short Form contains 10 of the 20 items of the original scale. It has good reliability and validity among individuals of Mexican descent and other ethnic groups (Grzywacz, Hovey, Seligman, Arcury, & Quandt, 2006). Even though the CESD and BDI have shown excellent psychometric properties across ethnic groups, these scales are lengthy and time consuming. Conversely, the PHQ-8 is shorter and can potentially be used on culturally diverse populations.

**Patient Health Questionnaire (PHQ)**

**Development**

The Patient Health Questionnaire emerged from the Primary Care Evaluation of Mental Disorder (PRIME-MD), which was created in the early 1990s. PRIME-MD evaluated many mental disorders, such as depression, anxiety, somatoform, substance abuse, and eating disorders (Kroenke et al., 2010). It contained 26 items with dichotomous (yes or no answers) responses. Even though it had shown good reliability
and validity (Spitzer et al., 1994), it had some limitations. First, it had too many items, so it affected the data collection process. With lengthy assessment, evaluators might not have time to assess other type of constructs (e.g., substance abuse), and participants take longer to return to their regular daily functions (e.g., work, homework). Second, it had to be administered by clinicians. This limitation affected the actual therapy session, given that a trained clinician had to execute the assessment. Thus, this type of administration might deviate from the actual goal of the session, which is providing therapy for the client. For these concerns, an alternative version of the PRIME-MD was developed by Spitzer and colleagues (1999), which reduced the limitations of the PRIME-MD, named the Patient Health Questionnaire (PHQ). The PHQ was self-administered, and shorter (16 items). The response choices range from 0 (not at all) to 3 (nearly every day). Spitzer and colleagues (1999) found that the PHQ and PRIME-MD had similar psychometric properties (e.g., reliability) as well sensitivity and specificity, yet the PHQ was viewed as better since it was the improved version of the PRIME-MD. One of the sub factors of the PHQ is comprised of the 9 items that measure depression, known as the 9-item Patient Health Questionnaire (PHQ-9). A positive characteristic is that the PHQ-9 is shorter. The PHQ (16 items) takes less than 8 minutes to complete and less than five minutes to score (Spitzer et al., 1999). This indicates that the PHQ-9 is probably faster to complete and score than its predecessors, such as the CES-D and BDI. Because the PHQ-9 has this characteristic, it has become widely used over the last few years for screening and research for depression (Kroenke et al., 2010).

**Different Versions**

The Patient Health Questionnaire (PHQ-9) emerged with some alternative forms,
such as the PHQ-8 (8 items), and the PHQ-2 (2 items). These PHQ-8 and PHQ-2 have shown good sensitivity and specificity in detecting individuals who are experiencing with mild to severe depression symptoms (Kroenke et al., 2010; Löwe, Kroenke, Herzog, & Gräfe, 2004). The PHQ-2 focuses on the items related to “sad mood” and “lack of desired” whereas the PHQ-8 does not include the 9\textsuperscript{th} item. The 9\textsuperscript{th} item is “Thoughts that you would be better off dead or hurting yourself in some way.” There are some reasons why researchers and clinicians do not utilize the 9\textsuperscript{th} item. One of the reasons is that this item contains a double-barreled item. The context of the item might create confusion among the clients since it is measuring two concepts simultaneously (Crede’ 2010). Then, there is the confusion from testers where they do not know which concepts the test-takers are really experiencing. Hence, the responses to the 9\textsuperscript{th} item do not indicate whether respondents are more suicidal or more hurting themselves because it lacks clarity. For this reason, the literature has recommended that double-barreled items must be treated with cautious (Crede’, 2010). Secondly, the PHQ-8 and the PHQ-9 are highly related (Kroenke et al., 2010; Razykov et al., 2012). Studies have documented that the PHQ-8 and the PHQ-9 had identical scoring thresholds for depression (Kroenke & Spitzer, 2002; Kroenke et al., 2009), indicating similar sensitivity and specificity. This suggests that the PHQ-8 and the PHQ-9 measure depression with the same criterion, so there is no surprise that researchers and clinicians are using the PHQ-8. Since the 9\textsuperscript{th} item lacks clarity and removing it does not jeopardize the PHQ-9’s psychometric properties, the 9\textsuperscript{th} item was not included in the current study.

**Psychometric Properties**

The Patient Health Questionnaire (PHQ-9) is a relatively new instrument. The
evaluation of its psychometric properties began in 1999 (Spitzer et al., 1999). The study utilized 3000 participants in order to evaluate the PHQ-9’s convergent and concurrent validity. The study found the PHQ-9 was highly related to the depression subscales from the Primary Care Evaluation of Mental Disorders, establishing concurrent validity. Additionally, the PHQ-9 was negatively related to well-being, establishing convergent validity. This original study showed that PHQ-9 had the potential to be used as an assessment tool. However, there are two important criticisms of this study. First, 79% of the sample was primarily European American. Thus, there was still insufficient evidence to support that the PHQ-9 might capture the same construct across ethnic groups. Second, the factor structure was not established. There was not any knowledge whether the PHQ-9 was unidimensional or multidimensional.

In attempts to fill these gaps, researchers have examined its factor structure and validity with some ethnic groups. Factor structure has been studied using a random sample of 5053 from the investigation of the PHQ in primary care and Obstetrics/Gynecology studies and Charles Wang Health Center in New York (Huang et al., 2006). The sample included Caucasians \( n = 2520 \), African Americans \( n = 598 \), Chinese Americans \( n = 941 \), and Latino \( n = 974 \) that were above the age of 18. Most of the sample included more women per ethnic group (more than 50%) than men. The Chinese Americans and Latinos completed the test in their native language. The exploratory factor analysis revealed that there was only one factor structure in each ethnic group. In addition, the internal reliability was good for Caucasians \( \alpha = .86 \), African Americans \( \alpha = .80 \), Chinese Americans \( \alpha = .79 \), and Latinos \( \alpha = .80 \). This study is important because it illustrates the factor structure. However, it only explored the factor
structure, and it did not test for factor invariance. Ryan, Bailey, Fearon, and King (2013) examined whether the PHQ-9 had factor invariance in face-to-face interviews or phone interviews. They collected self-reported data from 23,672 English speakers from the city of London, England, ranging from 16 to 112 years of age. They found that one factor was an excellent fit, and the PHQ-9 was able to meet the factor invariance across face-to-face and phone interviews. The unidimensional structure of the PHQ-9 was found across Surinam Dutch and Dutch citizens (Baas et al., 2011), as well as Asian men and women across different age groups (Yu, Tam, Wong, Lam, & Stewart, 2012).

For participants of Mexican descent, there is only one available study. Donlan and Lee (2010) were interested in examining the factor invariance of the Spanish version of the PHQ-9 among 123 immigrant farm workers of Mexican descent from the Health Outreach project in Oregon. Most of the participants (mean age = 30.7 years) did not finish elementary school and were living in the United States for at least 5 years. Since the participants had issues in reading and writing, the researchers read the survey while they self-reported. Utilizing confirmatory factor analysis, they found that the one factor structure was good, having excellent internal reliability ($\alpha = .91$). In addition, the PHQ-9 met convergent validity since it significantly related to self-esteem and quality of life. The researchers concluded that this was an excellent tool for measuring depressive symptoms among Mexican immigrant farm workers. However, there were some issues to address. First, some researchers recommend a sample of 200 or more, since it yields more stable results in SEM models (e.g., Kline, 2011; Tabachnick & Fidell, 2012), otherwise results should be interpreted with caution. The Donlan and Lee study had a sample size below 200; thus, its results might be questionable. Second, 93% percent of the sample
spoke native indigenous languages as well as Spanish, which means language and interpretation might impact the results. Thus, the results must be interpreted with cautious. Nevertheless, this investigation contributed to the literature on this topic, suggesting the possibility that a one factor structure might be the structure of the PHQ-9 for individuals of Mexican descent.

Additionally, investigators examined the factor invariance of the Spanish and English version of the PHQ-9 among 245 English and 234 Spanish speakers Latinos recruited from community sites (e.g., parks) in the San Diego County (Merz et al., 2011). In the Spanish speakers group, most of the sample did not finish high school, and 87% were born in Mexico. In the English speakers group, 60% reported being born in the United States. In both groups, there were only women, ranging from 18 to 30 years of age, who received 20 dollars for their participation. They found that the PHQ-9 Spanish ($\alpha = .85$) and English ($\alpha = .84$) versions had good internal consistency reliability as well as a one factor structure. In the multi-group confirmatory factor analysis, factor invariance (e.g., same structure, loadings, and latent mean) was met across ethnic groups. Yet, this studied only targeted women, and it did not examine the PHQ-9’s convergent validity. Nonetheless, the study provided more support of the PHQ-9’s psychometric properties and factor structure among Latinos who are Spanish and English speakers.

However, the unidimensional structure of the PHQ-9 is not universal in the available literature. To provide an example, researchers investigated the factor structure using 172 clients who were identified as experiencing high symptoms of depression from South Wales and Sydney, Australia (Titov et al., 2011). In their sample, there were more women (75%) than men, with a mean age of 44.7 years, ranging from 19 to 85 years.
Utilizing a confirmatory factor analysis, the study found a one factor structure did not fit well, but they did not provide an alternative model. This revealed that a one factor structure was not universal. Hence, researchers began exploring the possibility of other factor models for this scale. For instance, Pressler et al. (2011) were interested in investigating the factor structure of the PHQ-8 among 249 stroke patients from different outpatient clinics in the Midwest. The participants had experienced some type heart disease and were above the age of 18 years. In face-to-face interviews with this population, it was discovered that the PHQ-8 items were driven by two factors: affective and somatic depressive symptomatology. In addition, each one of the factors showed good internal reliability. Yet, the aforementioned researchers did not assess the tool’s validity.

Among Latinos, the two factor structure is also supported in the available literature. For instance, there is an investigation of the factor invariance among 16754 women (18 to 31 years of age) who agreed to participate after receiving an email throughout the United States (Granillo, 2012). More than 50% of the sample self-reported to be undergraduates; 1455 were Latinas and 15299 were White. The study found that an exploratory factor analysis revealed the possible existence of the two factor structure, showing excellent internal consistency reliability. In addition, the multi-group confirmatory factor analysis revealed that the PHQ-9’s two factor structure was invariant between the ethnic groups. In addition, the two factor structure met convergent construct validity since it was significantly related to mental health. The author concluded that the PHQ-9 might capture the multidimensionality of depression, yet still recommended the usage of the one factor structure. This study has some limitations. First, the Spanish
speaking participants were placed under Latinos as one group, ignoring how diversity of the Spanish speaking ethnic groups might impact the results. Second, the researcher removed two of the items in order to make the two factor structure fit. Third, men were not included. Finally, the researchers did not establish construct validity. Nevertheless, there is the possibility that the instrument might measure depression as multidimensional among Spanish speaking individuals.
CHAPTER III
METHODOLOGY

Procedures and Sample

An online and paper-pencil survey was developed and administered to research participants. The survey contained informed consent, questionnaire for demographics, and measures described below. This study was approved by the university’s Institutional Review Board (IRB).

The participants were recruited from the psychology subject pool. These group of students mainly came from lower division classes (e.g., introductory psychology course), and they participated to get extra credit for these courses. For the online survey version, the participants signed up for extra credit, and then they received the link for the survey. For the paper-pencil survey, trained research assistants (RAs) distributed the survey to students in their classrooms. After agreeing to participate, the participants completed the survey. It took them 12-16 minutes to complete the survey.

Participants who were not of Mexican descent were excluded for this study. The current sample included men and women who self-reported as individuals of Mexican descent, ranging from 18 to 29 years of age. Table 2 (attached in Appendix) provides the characteristics of the samples.

Measurement

A protocol was developed that included a section for demographics questionnaire that included questions regarding gender, age, ethnicity, generation status, education level, participant’s marital status, and parents’ marital statuses. These following paragraphs explained the measurements’ characteristics. Table 3 (attached in Appendix)
summarizes the excellent psychometric properties of the scales in the study.

**Patient Health Questionnaire (PHQ-8).** Emerging adults’ depressive symptoms were measured with the 8-item PHQ (Spitzer et al., 1999). A sample item follows: “Trouble falling asleep, staying asleep, or sleeping too much.” The response choices follow: 0 = not at all, 1 = several days, 2 = more than half the days, and 3 = nearly every day. For hypothesis 1 and hypothesis 2, each individual score was used for the analyses. For hypothesis 3, the items were averaged to create an overall score.

**The Center for Epidemiological Scale – Depression (CES-D) Short Form.** Emerging adults’ depressive symptoms were also assessed with the 10-item version (Kohout, Berkman, Evans, & Cornoni-Huntley, 1993) of the 20-item CES-D (Radloff, 1977). The items contained statements such as “I could not get ‘going’” and “I felt sad.” The response choices ranged from: 0 = rarely or none of the time (less than 1 day) to 3 = mostly or almost all the time (5 to 7 days). The CES-D had been used in research and clinical assessment in order to measure depression (Crockett et al., 2005). Multiple studies have documented how the CES-D has excellent psychometrical properties (e.g., reliability, validity, factor invariance) across diverse ethnic groups and age groups (Crockett et al., 2005; Nguyen et al., 2004; Radloff, 1977). The 10-item version was shown to have good factor structure and reliability in a sample of Mexican American migrant and immigrant adults (Grzywacz et al., 2006). Two items were reverse-coded, and then the ten items were averaged to create an overall score.

**Parental support.** Emerging adults’ perceptions of parents’ support (e.g., accepting, warm, and nurturing) were evaluated using the 4-item maternal and paternal subscales of the Parental Behavior Measure (Bush, Peterson, Cobas, & Supple, 2002). A
sample item follows, “This parent has made me feel that he/she would be there for me if I needed him.” Each item was asked twice, assessing the ratings of support for mother and then for father. The Likert scale ranged from $1 = \text{strongly disagree}$ to $4 = \text{strongly agree}$.

This instrument has shown good internal reliability among diverse ethnic groups (Behnke, Plunkett, Sands, & Bámaca-Colbert, 2011; Bush et al., 2002; Supple, Peterson, & Bush, 2004). All the items for mother and father were averaged to develop a subscale composite for parental support.

**Perceived stress.** Emerging adults’ perceptions of uncontrollable, unpredictable, and overloading life events were measured utilizing the 10-item Perceived Stress Scale (Cohen, Kamarck, & Mermelstein, 1983; Cohen & Williamson, 1988). Some sample items follow: “Been upset because of something that happened unexpectedly?” and “Felt nervous and ‘stressed’?” The response choices follow: $0 = \text{never}$, $1 = \text{almost never}$, $2 = \text{sometimes}$, $3 = \text{fairly often}$, and $4 = \text{very often}$. Four out 10 items were reversed coded. Researchers have documented the excellent reliability and validity of this tool with different ethnic groups (Flores et al., 2008; Lee, 2012; Roberti, Harrington, & Storch, 2006). All the items were averaged to create an overall score.

**Quality of life.** Emerging adult’s overall quality of life was measured with a single-item indicator (Zimmerman et al., 2006). Participants were asked “In general, how would you rate your overall quality of life during the last six months?” The response choices follow: $0 = \text{Very bad, my life could hardly be worse}$, $1 = \text{Pretty bad, most things are going poorly}$, $2 = \text{The good and bad parts are about equal}$, $3 = \text{Pretty good, most things are going well}$, and $4 = \text{Very good, my life could hardly be better}$. Zimmerman et al., (2006) reported high test-retest reliability (intraclass $r = .81$) and good convergent
validity. Specifically, depressed patients reported significantly lower quality of life than non depressed patients, and quality of life was significantly related to a range of theoretically related variables (Zimmerman et al., 2006).

Statistical Analyses

Hypothesis 1

A confirmatory factor analysis (CFA) was used for evaluating the PHQ-8’s factor structure for men and women of Mexican descent since the literature contains mixed findings about the factor structure (unidimensional and multidimensional) of this inventory. EQS software (Bentler, 2004) was utilized for the CFA. First, the model fit was examined. The PHQ-8 items were categorical, so the model fit was assessed using the Satorra-Bentler Scaled $\chi^2$, the Comparative Fit Index (CFI ≥ .95), and the Root Mean Squared Error of Approximation (RMSEA ≤ .06). These cutoffs were recommended in the SEM literature (Hooper, Coughlan, & Mullen, 2008; Hu & Bentler 1998, 1999; Kline, 2011; Tabachnick & Fidell, 2012). Meeting these criteria indicated acceptable fit. Additionally, the Lagrange multiplier test indicated how to modify the model in order to improve the fit. Second, the current study compared the factor structure models. The Akaike Information Criterion (AIC), a comparative measure of fit, was utilized in order to evaluate the factor models. When comparing models, researchers recommended that AIC values that are smaller indicated a better fit (Kline, 2011; Tabachnick & Fidell, 2012), indicating whether a one factor, two factor, or bi-factor structure was a better fit.

Hypothesis 2

After establishing the factor structure, a multi-group confirmatory factor analysis examined the factor invariance of the PHQ-8 between men and women of Mexican
descent. The model fit and software had the equivalent criteria as hypothesis 1. In addition, equality constraints were applied, forcing the parameters to be the same across ethnic groups. The Lagrange multiplier was used as post hoc, evaluating whether any of the constraints were significantly different or not.

To test the factor invariance between the target groups, the following steps took place. First, the proposed study evaluated configural invariance, which was the same factor structure. This indicated that fixed factor loadings, free factor loadings, and other parameters were the same across these ethnic groups. Second, the proposed study evaluated weak factorial invariance where equality constraints were applied on the loadings. Third, the proposed study examined the strict factorial invariance where equality constraints were applied on the unique variance and latent variable’s variance. Finally, the proposed study examined the equal factor means where equality constraints were applied as well.

**Hypothesis 3**

Next, the PHQ-8’s validity was evaluated for men and women of Mexican descent. For this aim, IBM’s SPSS statistics software was used. Additionally, a Pearson correlation tested the association between the PHQ-8, CES-D, and theoretically related variables. The following criteria was used to test the strength of the association: high correlation (a correlation between ± .50 and ± 1.00), moderate correlation (a correlation between ± .30 and ± .49), and small correlation (a correlation below ± .29).

First, the current study examined the concurrent criterion validity. A Pearson correlation tested the relationship between the PHQ-8 and the Center for Epidemiological Studies - Depression Scale (CES-D). Second, the proposed study examined the
convergent criterion validity utilizing perceived stress, parental support, and quality of life as variables. These variables’ relationships with depression have been well-documented in the literature (e.g., Husson et al., 2012; Ng & Jeffery, 2003). A Pearson correlation tested the relationship of the theoretically related variables and the PHQ-8.
CHAPTER IV
RESULTS

In the current study, there were no missing values. The results section was broken up into confirmatory factor analysis, concurrent validity, and then convergent validity of the PHQ-8.

Confirmatory Factor Analysis of the PHQ-8

A confirmatory factor analysis was used to test the model fit of the factor structure between men and women. The polychoric correlation matrix, mean, and standard deviation between the PHQ-8 items in the current study are presented in Table 4 (see Appendix). Table 5 (see Appendix) shows the $\chi^2$ changes and model modifications in the models for men and women.

For men, the one factor model’s fit was good: Satorra-Bentler Scaled $\chi^2$ (20) = 35.776, $p < .05$, CFI = .99, RMSEA = .063, and AIC = -4.223. The Lagrange Multiplier test suggested adding the correlation between errors in order to improve the model fit. Table 5 shows the $\chi^2$ changes and model modifications. In model 2, adding the association between the errors of items 6 and 4 was a significant improvement: Satorra-Bentler Scaled $\chi^2$ (19) = 31.483, $p < .05$, CFI = .99, RMSEA = .058, and AIC = -6.517. In the final model, adding the association between the errors of items 1 and 6 was a significant improvement: $\chi^2$ (18) = 24.996, $p > .05$, CFI = .996, RMSEA = .044, and AIC = -11.004. This was the final contribution to the model fit. In addition, the two factor (i.e., affective and somatic) model’s fit was excellent: Satorra-Bentler Scaled $\chi^2$ (19) = 28.080, $p > .05$, CFI = .99, RMSEA = .049, and AIC = -9.920. Since the model fit was excellent for these models, there was not any need to improve the fit. The latent variables
of affective and somatic were significantly and highly correlated, $r(202) = .92$, $p < .05$. Lastly, the bi-factor model had an excellent fit as well, Satorra-Bentler Scaled $\chi^2 (13) = 18.958$, $p > .05$, CFI = .991, RMSEA = .048, and AIC= -7.042. Suggestions by the Wald test were used to improve the model fit via dropping paths. In the final model, the path between items 8 and the sub factor of affective were dropped to significantly improve the fit: Satorra-Bentler Scaled $\chi^2 (14) = 18.850$, $p > .05$, CFI = .997, RMSEA = .041, and AIC = -9.150. After comparing the AIC’s values, the one factor model was lower than the two factor and the bi-factor models, which indicated that the one factor was a better model.

For women, the one factor model’s fit was acceptable: Satorra-Bentler Scaled $\chi^2 (20) = 53.920$, $p < .001$, CFI = .97, RMSEA = .054, and AIC = 13.920. In order to improve the fit, the Lagrange Multiplier test recommended adding the correlation between errors of items. In model 2, adding the association between the errors of items 2 and 4 was a significant improvement: Satorra-Bentler Scaled $\chi^2 (19) = 45.311$, $p < .001$, CFI = .976, RMSEA = .049, and AIC = 7.311. In the final model, adding the association between the errors of items 1 and 6 was a significant improvement: $\chi^2 (18) = 38.155$, $p > .001$, CFI = .981, RMSEA = .044, and AIC = 2.155. This was the final contribution to the model fit. In addition, the two factor (i.e., affective and somatic) model’s fit was excellent: Satorra-Bentler Scaled $\chi^2 (19) = 40.790$, $p > .05$, CFI = .98, RMSEA = .044, and AIC = 2.790. The model fit was acceptable for these models, improving the fit was not necessary. The latent variables of affective and somatic were significantly and highly correlated, $r(582) = .91$, $p < .05$. Lastly, the bi-factor model had an excellent fit, Satorra-Bentler Scaled $\chi^2 (14) = 14.920$, $p > .05$, CFI = .99, RMSEA = .030, and AIC = -6.077.
The bi-factor had a lower AIC than the other two models, indicating that the bi-factor model explained the data better for women. Also, item 8 and item 7 displayed small relationships with their correspondent factors, and the Wald test recommended dropping these paths in the following order. In model 2, dropping the path between item 8 and the sub factor of affective significantly improved the fit: Satorra-Bentler Scaled $\chi^2 (14) = 18.850$, $p > .05$, CFI = .999, RMSEA = .010, and AIC = -13.200. In the final model, dropping the path between item 7 and the sub factor of somatic significantly improved the fit: Satorra-Bentler Scaled $\chi^2 (15) = 15.444$, $p > .05$, CFI = 1.000, RMSEA = .007, and AIC = -14.556. Table 6 (see Appendix) summarizes the factor loadings in the current study. In the bi-factor for women, the loadings of the sub factors are weaker than the general factor, indicating that the PHQ-8 items are mainly driven by the depression latent factor.

Overall, for men, the one factor model was a better fit than the tested models, whereas for women the bi-factor model was a better fit. Because the PHQ-8’s structure was different for men and women, the factor invariance would not be tested. The one factor and bi-factor model are very applicable in a clinical, academic, and research setting for the following reasons: (1) the general factor of depression is mainly driven by all the items for men and women, and (2) clinicians and researchers typically utilize the one general score to screen for depression.

**Concurrent Validity of the PHQ-8**

Among men, depression (as measure in the PHQ-8) was significantly correlated with the previously validated CES-D short version, $r (112) = .61$, $p < .05$, with 37% shared variance. For the women, depression (as measured in the PHQ-8) was
significantly correlated with the previously validated CES-D short version, \( r (305) = .69, \)
\( p < .05, \) with 48\% shared variance. Overall, the PHQ-8 was strongly related to a well-
established depression inventory among men and women. Hence, the current study
established the PHQ-8’s concurrent validity.

**Convergent Validity of the PHQ-8**

Pearson correlations were used to examine the relationship between the
theoretically related variables and the PHQ-8 (see Table 7 in Appendix). In the current
study for both men and women, the PHQ-8’s depression factor was significantly and
positively related to perceived stress. Also, the PHQ-8’s depression was significantly and
negatively related to parental support and quality of life. Overall, the PHQ-8 was related
with the theoretically related variables in the hypothesized directions among men and
women. Hence, the current study established the PHQ-8’s convergent validity.
CHAPTER V
DISCUSSION

The purpose of this study was to evaluate whether the PHQ-8 is a reliable and valid scale for emerging adult men and women of Mexican descent. This population was selected for the following reasons: (1) because of significant life changes and transitions, emerging adults are at risk of psychopathology; (2) typically, individuals of Mexican descent report higher rates of depressive symptoms than other ethnic groups; and (3) there is scarcity of research on evaluating mainstream depression instruments’ psychometric properties for emerging adult men and women of Mexican descent. In addition, there is a need for a quick and DSM-5 related screener, given that many of the widely used depression scales are lengthy and deviate from the DSM-5’s depression criteria. From a practical perspective, a short, reliable, and valid depression scale would be most favorable in order to measure depression symptomatology, considering that depression is a major health concern worldwide (Lopez & Mathers, 2006; CDC, 2010), and even in college settings (American College Health Association, 2008; Del Pilar, 2008; Furr, McConnell, Westefeld, & Jenkins, 2001; Voelker, 2003).

Summary of Findings and Discussion

In the current study, results indicated good fit for a one factor solution for men. Researchers have documented that the PHQ-8 has one factor structure among Latinos/as (Huang et al., 2006; Merz et al., 2011), as well among men of Mexican descent (Donlan & Lee, 2010). The current study’s findings support the PHQ-8’s unidimensional structure for young men of Mexican descent. However, results indicated a good fit for a bi-factor solution for young women. To our knowledge, this is the first time that the PHQ-8 has
been studied for its potential bi-factor model, and there is not any literature on this topic. Thus, the present investigation contributes significantly to the field of psychometrics.

There are three major issues to highlight about this finding. First, the items are still driven by the depression factor. Second, there are still sub factors present in the PHQ-8, which recognizes its multidimensionality. However, the PHQ-8’s depression factor had stronger loadings than the sub factors. Thus, the PHQ-8 is mainly one-dimensional. Having one dimension for this scale is important because most researchers and practitioners are likely to use the PHQ-8 composite score (i.e., average or sum of all items) for the quantification of depressive symptoms. Moreover, there are important concepts to highlight about the PHQ-8’s factor models tested in this study. The one factor, two factor, and bi-factor all had great fit. Researchers and theorists who are interested in studying any of these models can continue researching them and applying them in theoretical frameworks.

In addition, in the current study it was not feasible to test the PHQ-8’s factor invariance between men and women of Mexican descent. Each one of these groups had different factor structures. The difference in the factor structures might be due to gender differences, both genetic and socialization-driven, in processing as well as experiencing depression for men and women. In the future, researchers and theorists should further investigate the PHQ-8’s factor structure across different ethnic groups and evaluate its factor invariance because the PHQ-8 lacks research across many ethnic groups, especially gender-wise, and is not adequately capturing depression within these groups.

Next, in the current study, the PHQ-8’s concurrent validity was established for emerging adult men and women of Mexican descent. The PHQ-8’s total score was
strongly related to the 10-item version of the CES-D. The current findings support previous research showing that the PHQ-8 and PHQ-9 are highly correlated with well-established scales of depression (e.g., Ganguly et al., 2013; Kroenke et al., 2010; Kroenke et al., 2009; Richardson et al., 2010). Conducting the present analysis was important to show that the PHQ-8 captures the same construct as other widely used and established depression inventories.

Finally, the PHQ-8’s convergent validity was established for young men and women of Mexican descent. The PHQ-8’s total score was significantly correlated with documented and theoretically related variables in the expected directions. In this regard, the PHQ-8 was positively associated to perceived stress, whereas it was negatively correlated to parental support and quality of life. The current findings support the results of prior studies that stress can exacerbate depressive symptoms (e.g., Flores et al., 2008; Ng & Jeffery, 2003; Vega et al., 2010); parental support declines depressive symptoms (e.g., Yeung, & Leadbeater, 2010; Liem et al., 2010; Needham, 2008); and depression is detrimental to quality of life (e.g., Abraham & Fava, 1999; Gotlib, 1997). Because scores on the PHQ-8 were related to theoretically-related variables in expected directions, the PHQ-8 should be a good indicator of depressive symptomatology for researchers studying predictors of depression as well as outcomes of depression among young men and women of Mexican descent.

**Limitations and Research Implications**

This thesis contributes successfully to filling the gap in evaluating depressive scales’ psychometric properties for use with young men and women of Mexican descent. However, there are certain limitations of the present investigation that need to be
addressed. First, the participants are university students in Southern California, and the majority of them are born in the United States. Thus, this limits the generalizability of the present findings. In future studies, researchers should examine the psychometric properties of the PHQ-8 in samples of non-college emerging adults of Mexican descent and other Latino/a emerging adults, as well as in other ethnic minority samples. Also, young adults of Mexican descent were collapsed into one group, which does not take into consideration differences by generational status or place of birth (e.g., U.S. born vs. not U.S. born). In future psychometric studies in this area, scholars should examine generational status or place of birth, because these factors might play an unexplored role in the findings. Also, the 9th item was not examined in the current study, based on recommendations on removing the doubled barrel items (Crede’, 2010). Thus, in future studies, researchers may want to examine the psychometric properties of the PHQ-9 within samples of emerging adults of Mexican descent.

Implication and Conclusion

In summary, the PHQ-8 produced similar results for emerging adult men and women of Mexican descent. Despite the fact that there was no evaluation of factor invariance, the PHQ-8 has excellent psychometric properties given that it is a primarily a unidimensional depression inventory. In this study, the concurrent and convergent validity of the PHQ-8 were established. Thus, this is a tool that can successfully quantify depressive symptomatology in emerging adult men and women of Mexican descent. More importantly, community mental health care practitioners, college counselors, medical professionals, as well as social scientists can utilize the PHQ-8 with the target population.
REFERENCES


Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modeling:


Needham, B. L. (2008). Reciprocal relationships between symptoms of depression and parental support during the transition from adolescence to young adulthood. *Journal of Youth and Adolescence, 37*(8), 893-905. doi:10.1007/s10964-007-9181-7


APPENDIX

Table 1  

<table>
<thead>
<tr>
<th>Item</th>
<th>PHQ-9</th>
<th>DSM-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Feeling down, depressed, irritable, or hopeless</td>
<td>Feeling depressed for the majority of the day</td>
</tr>
<tr>
<td>2</td>
<td>Little interest or pleasure in doing things</td>
<td>Significant lack of interest or pleasure in most activities almost all day</td>
</tr>
<tr>
<td>3</td>
<td>Trouble falling asleep, staying asleep, or sleeping too much</td>
<td>Significant change in sleep patterns, such as insomnia or hypersomnia</td>
</tr>
<tr>
<td>4</td>
<td>Poor appetite, weight loss, overeating</td>
<td>Marked weight loss that cannot be attributed to dieting, excessive weight gain, or increased appetite</td>
</tr>
<tr>
<td>5</td>
<td>Feeling tired, or having little energy</td>
<td>Feeling tired or fatigued</td>
</tr>
<tr>
<td>6</td>
<td>Feeling bad about yourself or feeling that you are a failure, or that you have let yourself or your family down</td>
<td>Feeling disproportionately guilty or worthless</td>
</tr>
<tr>
<td>7</td>
<td>Trouble concentrating on things like schoolwork, reading, or watching TV</td>
<td>Difficulty in executive functioning (e.g. being indecisive or unable to concentrate)</td>
</tr>
<tr>
<td>8</td>
<td>Moving or speaking so slowly that other people could have noticed. Or the opposite – being so fidgety or restless that you were moving around a lot more than usual</td>
<td>Physical agitation or retardation</td>
</tr>
<tr>
<td>9</td>
<td>Thoughts that you would be better off dead of or hurting yourself in some way</td>
<td>Persistent thoughts of death or suicide without a specific plan, or attempted suicide, or a specific suicide plan</td>
</tr>
</tbody>
</table>
Table 2  
Demographic Characteristics of the Sample

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N = 804)</td>
<td>(n = 600)</td>
<td>(n = 204)</td>
<td></td>
</tr>
<tr>
<td><strong>Age (Range is 18-29 years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>19.2</td>
<td>19.2</td>
<td>19.3</td>
</tr>
<tr>
<td>SD</td>
<td>1.6</td>
<td>1.6</td>
<td>1.6</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>25.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>75.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>U.S Born</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>91.2%</td>
<td>90.9%</td>
<td>92.2%</td>
</tr>
<tr>
<td>No</td>
<td>8.8%</td>
<td>9.2%</td>
<td>7.8%</td>
</tr>
<tr>
<td><strong>Parents’ Marital Status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>69.8%</td>
<td>70.2%</td>
<td>68.6%</td>
</tr>
<tr>
<td>Divorced</td>
<td>9.2%</td>
<td>8.8%</td>
<td>10.3%</td>
</tr>
<tr>
<td>Never married</td>
<td>8.1%</td>
<td>8.3%</td>
<td>7.4%</td>
</tr>
<tr>
<td>Remarried</td>
<td>2.1%</td>
<td>2.0%</td>
<td>2.5%</td>
</tr>
<tr>
<td>Widowed</td>
<td>4.2%</td>
<td>4.5%</td>
<td>3.4%</td>
</tr>
<tr>
<td>Separated</td>
<td>5.0%</td>
<td>4.7%</td>
<td>5.9%</td>
</tr>
<tr>
<td>Other</td>
<td>1.5%</td>
<td>1.5%</td>
<td>1.5%</td>
</tr>
<tr>
<td>Missing</td>
<td>0.1%</td>
<td>0.0%</td>
<td>0.5%</td>
</tr>
<tr>
<td><strong>Reside with Parent</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>70.0%</td>
<td>70.7%</td>
<td>68.1%</td>
</tr>
<tr>
<td>No</td>
<td>29.9%</td>
<td>28.7%</td>
<td>29.4%</td>
</tr>
<tr>
<td>Missing</td>
<td>0.1%</td>
<td>0.0%</td>
<td>0.5%</td>
</tr>
<tr>
<td><strong>Classification</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freshman</td>
<td>53.0%</td>
<td>52.7%</td>
<td>53.9%</td>
</tr>
<tr>
<td>Sophomore</td>
<td>27.2%</td>
<td>27.7%</td>
<td>26.0%</td>
</tr>
<tr>
<td>Junior</td>
<td>13.2%</td>
<td>13.3%</td>
<td>12.7%</td>
</tr>
<tr>
<td>Senior</td>
<td>6.2%</td>
<td>6.0%</td>
<td>6.9%</td>
</tr>
<tr>
<td>Masters</td>
<td>0.2%</td>
<td>0.3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Missing</td>
<td>0.1%</td>
<td>0.0%</td>
<td>0.5%</td>
</tr>
<tr>
<td><strong>Generation Status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st generation (all foreign born)</td>
<td>9.8%</td>
<td>10.0%</td>
<td>9.3%</td>
</tr>
<tr>
<td>2nd generation (subject USA born, parents foreign born)</td>
<td>77.9%</td>
<td>77.3%</td>
<td>79.4%</td>
</tr>
<tr>
<td>2.5 generation</td>
<td>10.0%</td>
<td>10.0%</td>
<td>9.8%</td>
</tr>
<tr>
<td>3rd generation (all USA born)</td>
<td>4.7%</td>
<td>2.5%</td>
<td>1.5%</td>
</tr>
<tr>
<td></td>
<td>All $(N = 804)$</td>
<td>Women $(N = 600)$</td>
<td>Men $(N = 204)$</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------------</td>
<td>------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>CESD</td>
<td>.84</td>
<td>.83</td>
<td>.85</td>
</tr>
<tr>
<td>Parental support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother support</td>
<td>.86</td>
<td>.86</td>
<td>.88</td>
</tr>
<tr>
<td>Father support</td>
<td>.89</td>
<td>.88</td>
<td>.89</td>
</tr>
<tr>
<td>Perceived stress</td>
<td>.82</td>
<td>.82</td>
<td>.82</td>
</tr>
<tr>
<td>PHQ</td>
<td>.88</td>
<td>.87</td>
<td>.89</td>
</tr>
</tbody>
</table>
### Table 4
Means, Standard Deviations and Polychoric Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Feeling down, depressed, irritable, or hopeless?</td>
<td>--</td>
<td>.75</td>
<td>.60</td>
<td>.54</td>
<td>.48</td>
<td>.83</td>
<td>.62</td>
<td>.53</td>
</tr>
<tr>
<td>2. Little interest or pleasure in doing things?</td>
<td>.72</td>
<td>--</td>
<td>.52</td>
<td>.47</td>
<td>.47</td>
<td>.63</td>
<td>.53</td>
<td>.48</td>
</tr>
<tr>
<td>3. Trouble falling asleep, staying asleep, or sleeping too much?</td>
<td>.53</td>
<td>.56</td>
<td>--</td>
<td>.70</td>
<td>.63</td>
<td>.63</td>
<td>.65</td>
<td>.49</td>
</tr>
<tr>
<td>4. Poor appetite, weight loss, or overeating?</td>
<td>.54</td>
<td>.51</td>
<td>.68</td>
<td>--</td>
<td>.63</td>
<td>.6</td>
<td>.70</td>
<td>.65</td>
</tr>
<tr>
<td>5. Feeling tired, or having little energy?</td>
<td>.54</td>
<td>.53</td>
<td>.62</td>
<td>.67</td>
<td>--</td>
<td>.54</td>
<td>.62</td>
<td>.6</td>
</tr>
<tr>
<td>6. Feeling bad about yourself – or feeling that you are a failure, or that you have let yourself or your family down?</td>
<td>.67</td>
<td>.58</td>
<td>.42</td>
<td>.45</td>
<td>.51</td>
<td>--</td>
<td>.76</td>
<td>.66</td>
</tr>
<tr>
<td>7. Trouble concentrating on things such as schoolwork, reading, or watching TV?</td>
<td>.6</td>
<td>.58</td>
<td>.54</td>
<td>.57</td>
<td>.59</td>
<td>.59</td>
<td>--</td>
<td>.57</td>
</tr>
<tr>
<td>8. Moving or speaking so slowly that other people could have noticed?...Or the opposite – being so fidgety or restless that you were moving around a lot more than usual?</td>
<td>.55</td>
<td>.5</td>
<td>.5</td>
<td>.49</td>
<td>.47</td>
<td>.54</td>
<td>.56</td>
<td>--</td>
</tr>
</tbody>
</table>

Men’s M

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men’s SD</td>
<td>.66</td>
<td>.73</td>
<td>.9</td>
<td>.56</td>
<td>1.07</td>
<td>.6</td>
<td>.98</td>
<td>.4</td>
</tr>
</tbody>
</table>

Women’s M

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women’s SD</td>
<td>.76</td>
<td>.75</td>
<td>1.1</td>
<td>.89</td>
<td>1.37</td>
<td>1.7</td>
<td>.47</td>
<td>1.05</td>
</tr>
</tbody>
</table>

Note: All correlations were significant at least $p < .05$. Numbers below -- are the women’s values and above -- are the men’s values.
Table 5

$\chi^2$ Changes to Improve the Fit of the PHQ-8’s Confirmatory Factor Analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>$\chi^2$Δ</th>
<th>CFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>35.78</td>
<td>.99</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>Model 2 Add correlation between the errors of items 6 and 4</td>
<td>31.48</td>
<td>4.3</td>
<td>.99</td>
<td>.06</td>
</tr>
<tr>
<td>Model 3 Add correlation between the errors of items 1 and 6</td>
<td>28.08</td>
<td>3.4</td>
<td>.99</td>
<td>.05</td>
</tr>
<tr>
<td>Bi-Factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>18.96</td>
<td>.99</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Model 2 Remove path from item 8 and factor 1</td>
<td>18.85</td>
<td>.11</td>
<td>.99</td>
<td>.04</td>
</tr>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>53.92</td>
<td>.97</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Model 2 Add correlation between the errors of items 2 and 4</td>
<td>45.31</td>
<td>8.61</td>
<td>.98</td>
<td>.05</td>
</tr>
<tr>
<td>Model 3 Add correlation between the errors of items 1 and 6</td>
<td>38.16</td>
<td>7.15</td>
<td>.98</td>
<td>.04</td>
</tr>
<tr>
<td>Bi-Factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>14.92</td>
<td>.99</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Model 2 Remove path between item 8 and factor 1</td>
<td>14.80</td>
<td>.12</td>
<td>.99</td>
<td>.01</td>
</tr>
<tr>
<td>Model 3 Remove path between item 7 and factor 2</td>
<td>15.44</td>
<td>.64</td>
<td>1.00</td>
<td>.01</td>
</tr>
</tbody>
</table>
Table 6
Factor Loadings of the Confirmatory Factor Analyses for Men and Women

<table>
<thead>
<tr>
<th>PHQ-items</th>
<th>One Factor</th>
<th>Two Factor</th>
<th>Bi-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Depression</td>
<td>Affective</td>
<td>Somatic</td>
</tr>
<tr>
<td>Men</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>.87*</td>
<td>.95*</td>
<td>.64*</td>
</tr>
<tr>
<td>Item 2</td>
<td>.76*</td>
<td>.85*</td>
<td>.73*</td>
</tr>
<tr>
<td>Item 3</td>
<td>.87*</td>
<td>.88*</td>
<td>.76*</td>
</tr>
<tr>
<td>Item 4</td>
<td>.91*</td>
<td>.91*</td>
<td>.80*</td>
</tr>
<tr>
<td>Item 5</td>
<td>.83*</td>
<td>.84*</td>
<td>.75*</td>
</tr>
<tr>
<td>Item 6</td>
<td>.98*</td>
<td>.98*</td>
<td>.95*</td>
</tr>
<tr>
<td>Item 7</td>
<td>.90*</td>
<td>.93*</td>
<td>.86*</td>
</tr>
<tr>
<td>Item 8</td>
<td>.84*</td>
<td>.88*</td>
<td>.87*</td>
</tr>
<tr>
<td>Women</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>.88*</td>
<td>.88*</td>
<td>.83*</td>
</tr>
<tr>
<td>Item 2</td>
<td>.84*</td>
<td>.84*</td>
<td>.76*</td>
</tr>
<tr>
<td>Item 3</td>
<td>.83*</td>
<td>.83*</td>
<td>.72*</td>
</tr>
<tr>
<td>Item 4</td>
<td>.85*</td>
<td>.85*</td>
<td>.71*</td>
</tr>
<tr>
<td>Item 5</td>
<td>.82*</td>
<td>.82*</td>
<td>.73*</td>
</tr>
<tr>
<td>Item 6</td>
<td>.80*</td>
<td>.80*</td>
<td>.74*</td>
</tr>
<tr>
<td>Item 7</td>
<td>.82*</td>
<td>.83*</td>
<td>.82*</td>
</tr>
<tr>
<td>Item 8</td>
<td>.74*</td>
<td>.74*</td>
<td>.73*</td>
</tr>
</tbody>
</table>

*p < .05. Notes: UnStnd = Unstandardized, SE = Standard error, Stnd = Standardized.
Table 7

*Pearson Correlations between PHQ and Theoretically-Related Variables.*

<table>
<thead>
<tr>
<th></th>
<th>Depression</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Men</strong></td>
<td></td>
</tr>
<tr>
<td>CESD (n = 114)</td>
<td>.61*</td>
</tr>
<tr>
<td>Quality of life (n = 204)</td>
<td>-.59*</td>
</tr>
<tr>
<td>Perceived Stress (n = 204)</td>
<td>.67*</td>
</tr>
<tr>
<td>Parental Support (n = 204)</td>
<td>-.27*</td>
</tr>
<tr>
<td><strong>Women</strong></td>
<td></td>
</tr>
<tr>
<td>CESD (n = 302)</td>
<td>.69*</td>
</tr>
<tr>
<td>Quality of life (n = 594)</td>
<td>-.51*</td>
</tr>
<tr>
<td>Perceived Stress (n = 594)</td>
<td>.67*</td>
</tr>
<tr>
<td>Parental Support (n = 594)</td>
<td>-.26*</td>
</tr>
</tbody>
</table>

*p < .05.*