MEASUREMENT INVARIANCE OF THE SF-12 ACROSS ETHNIC GROUPS AMONG WOMEN IN POSTPARTUM

by

TAMER FAROUK DESOUKY

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN PSYCHOLOGY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2011

DEDICATION

بسم الله الرحمن الرحيم

In the Name of Allah, the Most Gracious, the Ever Merciful

I'd like to take the time to recognize and appreciate my parents, Farouk Desouky and Aisha Shafyie, who have given everything for my education and happiness. I love you two very dearly. Words truly cannot express what I feel in my heart for the both of you. May Allah bless you both and give you *Al-Firdous Al-'Alaa;* Allahumma Ameen, Allahumma Ameen, Allahumma Ameen.

ACKNOWLEDGEMENTS

With much gratitude and appreciation, I thank Dr. Pablo Mora for his patience and assistance with my thesis project. Thank you again for your guidance.

Dr. Andrew Baum deserves a special acknowledgement. Although I constantly barraged him with questions and shared with him my various life goals (and by extension my various existential crises) he was always patient with me and made a joke out of any difficult situation. Andy, you truly left an impression on us and thank you for always being there when we needed it. You will be missed.

To my fellow graduate student colleagues I thank you all for the moral support and the “keep this on the down low” advice I received throughout this whole process! Special thanks to Bilal Khan, Torry Dennis, Deborah McArthur, Steven Diamond, Jason Popan, and Lara Trevino.

Finally, to my lovely “kitabi” wife, Amal Saadat. My best friend, thank you for all the laughs and your encouraging smile.

April 14, 2011

ABSTRACT

MEASUREMENT INVARIANCE OF THE SF-12 ACROSS ETHNIC GROUPS AMONG WOMEN IN POSTPARTUM

Tamer F. Desouky, M.S.

The University of Texas at Arlington, 2011

Supervising Professor: Pablo A. Mora

Purpose: The purpose of this study was to examine differential item functioning (DIF) as a plausible cause of subgroup variation in the short-form health survey (SF-12). *Methods*: Cross sectional secondary data analysis of postpartum women (*n* = 655) who participated in a prospective cohort longitudinal design was conducted. Hypotheses predicted that subgroups would exhibit DIF among items measuring physical health as assessed by the SF-12. If DIF is found then other postpartum related variables, such as social support and parity, will explain DIF through mediation analysis. A Multiple Indicator Multiple Cause (MIMIC) model was used to examine DIF among these subgroups of women. *Results*: Items SF1 "self assessed general health", SF8 "bodily pain", and SF9 "calm and peaceful" all indicated DIF. However, only DIF effects of African-Americans endorsing SF8 (OR = 2.11, CI_{95} = 1.20, 3.71) and Hispanics endorsing SF9 (OR = 2.62, CI_{95} = 1.64, 4.17) signified meaningful effect sizes as indicated by the Odds-Ratio values. Further examination of the SF1 "self assessed general health" and SF8 "bodily pain" DIF effects, relevant to the hypotheses of this study, was conducted in order to explain the DIF effect. The DIF effect of Hispanics endorsing item SF1 "self assessed general health" more than Caucasians explained only 47% of the total effect

present, while on the other hand, the DIF effect for African-Americans endorsing the SF1 item explained 21% of the variance in the total effect. Supplementary assessment using social support as a mediator in the SF1 “self assessed general health” DIF effect revealed a significant relationship. Specifically, social support partially mediated the DIF effect for both ethnicities. Thus, social support explained the differential response of both of these ethnicities to this specific item. The SF8 “bodily pain” item had a meaningful DIF effect, and for Hispanics explained 32% while for African-Americans accounted for 50% of the variance in the total effect. Employing both social support and parity as mediators yielded a significant relationship. Both mediators partially mediated the DIF effect for African-Americans but only social support partially mediated the DIF effect for Hispanics. *Conclusion*: The results of this study reveal items SF8 “bodily pain” and SF9 “calm and peaceful” as biased towards Hispanics and African-Americans, respectively, after matching on overall mental or physical health, compared to Caucasian women. However, utilizing mediation analysis explained the DIF effects and provided more understanding of the bias towards these ethnicities. Implications of this study are to use more than one method to assess DIF and to use psychological theory to explain the DIF effects through mediation analysis.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction & Aims

The purpose of this study was to ascertain the validity of the short-form health survey (SF-12), a commonly used health-related quality of life (HRQoL) measure. HRQoL indicators provide important information about adjustment to illness and treatment efficacy in various populations. Exclusive of social or emotional functioning, physical functioning is an essential component of HRQoL and measurement of this dimension provides insight into domains not directly observable, such as the impact of illness burden, disability, physical impairment, and pain on well-being. However, HRQoL indicators must be validated for accurate measurement of all populations, especially minority populations.

Aim 1. The first aim of this study was to determine whether the SF-12 is invariant across women of different ethnic groups in postpartum (i.e., Caucasians, African-Americans, and Latinas). The focus of this study was to determine whether the various ethnic groups differentially endorse items of physical health (i.e., differential item functioning or DIF).

Aim 2. The second aim of this study was to explain subgroup variation (i.e., DIF) in the SF-12 by moving beyond typical sociodemographic differences and examining whether specific postpartum variables, such as social support and parity, explain DIF.

1.2 Measurement and Health Disparities

Healthcare access and utilization of healthcare services differ markedly across racial, ethnic, and socioeconomic groups in the United States despite technological advancement in medical services and improvement in the delivery of these health services (Cook, McGuire, & Miranda, 2007; Dobalian & Rivers, 2008).

Lack of standardization of various research related terminologies (e.g., race vs. ethnicity) and inconsistent data collection procedures and instruments contribute to the lack of understanding of the factors that underlie healthcare inequalities (McDonough et al., 2004; National Academy of Sciences, 2009). In addition, self-report instruments used to ascertain mental health in survey studies may not be equally valid across different ethnic groups. Due to poor instrumentation, items on a scale could be biased against a minority group and result in spurious scores. In addition, if the validity of these items is disproportionately affected by subgroup characteristics, such as cultural, linguistic, generational, or economical, then subsequent diagnosis and treatment for these groups could become ineffectual and/or problematic because they rely on inaccurate assessment of health conditions. Despite the widespread use of self reports in behavioral sciences, scales are rarely tested to determine if metric properties are consistent across various groups (e.g., ethnic, age, or sex).

In addition to ensuring that assessment is unbiased, the examination of DIF can help improve the understanding of psychological phenomena. For example, in a study examining the invariance of the Diagnostic Interview Schedule (DIS) used to measure depression; Gallo, Anthony and Muthén (1994) found that older adults (65 years and above) were less likely to endorse an item measuring dysphoria (a key facet of depressive symptomatology). This finding has two important implications. On the one hand, older adults' tendency not to endorse dysphoria items could account for the low rates of Major Depression in this age group. On the other hand, this evidence suggests that older adults experience depression differently than younger adults. Gallo and Rabins (1999) confirmed and extended these findings by showing that older adults were also less likely to endorse items assessing anhedonia. Together, these data suggest that older adults may have significant clinical depression but are under diagnosed because they do not meet required criteria due to instrumentation bias against their specific age group. Under recognition can result in the under treatment of depressive symptoms among the

elderly which can lead to decreased quality of life. This example highlights the problems that can arise from improper measures and how they could, potentially, lead to health disparities.

1.3 Measurement Invariance

In measurement theory, each instrument is made up of numerous items that represent different subscales or facets, which in return represent different factors. Responses to items are observed and are used as indicators of an unobserved variable (also called the latent variable). Because the latent variable is not measured directly, it is inferred indirectly through its presumed effect on the observed responses to the scale items (Fleishman, Spector, & Altman, 2002). Therefore, the individual's response to an item reflects a causal effect of the latent variable.

A key assumption of psychometrics is that individuals' characteristics unrelated to the construct do not affect the way they respond to the items (i.e., the measure is invariant). This means that the questionnaire measures the same construct in the same way in different independent groups (Williams et al., 2007). However, this assumption of invariance or equivalence must be tested. One way to measure invariance is by identifying the presence or absence of DIF. DIF is also referred to by other terms such as "item bias, item-response bias, measurement noninvariance, measure bias, measure disturbance, test-irrelevant variance, or factorial invariance" (Jones & Gallo, 2002, pp. 549). DIF is present when two or more groups that are matched or equal on the trait being assessed, have different probabilities of responding to or endorsing a given item (Mazor, Hambleton, & Clauser, 1998; Finch & French, 2008; Teresi, Kleinman, & Ocepek-Welikson, 2000). The presence of DIF indicates that a different variable, other than the latent variable (e.g., different interpretation of the item), is influencing performance on that given item (Mazor, Hambleton, & Clauser, 1998; Finch & French, 2008; Fleishman & Lawrence, 2003). Lack of measurement invariance indicates that items lack validity for one of the groups and confounds group comparisons (Steinberg & Thissen, 2006). Additionally, item-level biases can carry over to factor-level biases and affect overall scale

structure (Cole, Kawachi, Maller, & Berkman, 2000). For a simplified model of DIF please refer to Figure 1.1.

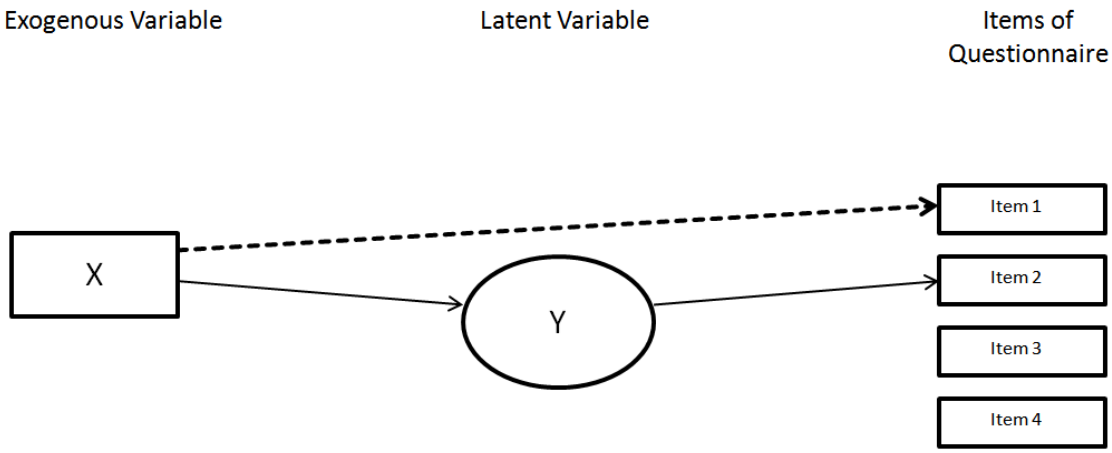

Figure 1.1 Simplified DIF Model.
Each arrow represents a causal relationship from one construct to the other. Indicators (Items) on the right are influenced by the latent variable. The dashed arrow from exogenous variable "X" to "Item 1" represents one possible DIF effect that is not mediated by the latent variable.

Conducting invariance analysis provides two benefits. First, from a measurement perspective, biased items can be modified or removed in order to provide a more precise and non-biased measure. As a result, the assessment of constructs based on these items will be accurate. Second, practical implications of invariance analysis include valid group comparisons and quite possibly a more precise indicator of minority health. For that reason, Teresi (2006) highlighted the importance of measuring item invariance in relation to health disparities.

CHAPTER 2

THE POSTPARTUM PERIOD

The postpartum period, which lasts approximately 6 months after birth, is a period that encompasses many physical and social changes. Typically, for the majority of women this period will pass without lasting changes to their body. However, for some women there may be long-term physical and/or psychological disability (Littleton & Engebretson, 2002; Alexander, Levy, & Roch, 1990). Therefore, it is important to understand the various physical limitations that may delay recovery or offset adjustment to postpartum.

2.1 Physical Changes

2.1.1 Physical Changes that Cause Pain

The natural and abnormal processes that cause pain are numerous. For example, during the childbirth process many complications can occur. For some women who must undergo a cesarean delivery the pain of stitches to the abdominal wall can last several months. Additionally, an episiotomy might be performed for some women undergoing vaginal delivery. These procedures and surgeries result in postoperative pain that can last for the duration of postpartum (Littleton & Engebretson, 2002). Moreover, abdominal and vaginal tears while giving birth can lead to muscle aches, with the most notable manifestation being backaches. The birth process may also cause hemorrhoids to develop which may cause pain during defecation, which if severe, can lead to another problem of constipation (Littleton & Engebretson, 2002; Klossner, 2006).

2.1.2 Physical Changes that Cause Discomfort

During this period, the body will regress back to its previous state before pregnancy and also accommodate to new unique situations that call for hormonal changes, such as breastfeeding (Littleton & Engebretson, 2002; Klossner, 2006).

These changes usually cause physical discomfort, which may be shown as irritability, mood swings, and general fatigue. For example, during this time the uterus contracts in parallel with oxytocin release which causes afterpains (Klossner, 2006). Uterine contraction leads to involution and eventually the uterus returns to normal size after 9-10 days (Littleton & Engebretson, 2002). Moreover, postpartum bleeding (lochia) is a normal process whereby the uterine wall is shed along with blood, mucus, and white blood cells (Klossner, 2006). However, if this becomes severe, then hemorrhaging can occur which can put the mother's life in danger. Also, during this period excess fluid, consisting of plasma, blood, and water are eliminated. For example, during childbirth, blood loss of 300-500 mL is normal for vaginal delivery and 500-1,000 mL for cesarean delivery (Littleton & Engebretson, 2002; Klossner, 2006). Also, excess plasma is eliminated through frequent urination. Finally, excessive sweating (diaphoresis) will get rid of extra fluid, especially during the night, explaining the phenomenon of night sweats (Klossner, 2006). Moreover, suckling by the infant stimulates prolactin release, which subsequently increases milk production (Klossner, 2006). This can lead to swollen or tender breasts and may be a discomfort for some women.

2.2 Socioemotional Changes

2.2.1 Marital Relationship

Socioemotional changes that occur postpartum can transpire in any of the following domains: relationships, sexuality, identity roles, and responsibilities. Several research studies provide interesting facts concerning the marital relationship. For example, Wallace and Gotlib (1990) found that women who planned their pregnancies tended to have poor marital adjustment at postpartum than woman who did not plan their pregnancies. Moreover, if infant temperament is difficult and unpredictable then poor marital adjustment was evident (Wallace & Gotlib, 1990). On the other hand, having good marital adjustment during pregnancy and positive perceived parenting competence predicts high postpartum marital adjustment (Belsky & Rovine, 1990). However, the trend does not hold over time. Marital adjustment tends to peak at one

month postpartum and steadily declines by six months due to numerous factors (Belsky & Rovine, 1990).

2.2.2 Identity Roles and Responsibilities

Identity roles change and responsibilities increase after childbirth. The postpartum mother must now be able to understand her new roles, adapt to the constant demands, and be able to successfully solve the various problems associated with infant care. Although marital adjustment may decline in postpartum because of the parenting stressors, social support plays a key role in alleviating the constant demands as well as previous parental experience in dealing with infant care.

Consequently, as will be described in the next chapter, it is of utmost importance to study physical health in relation to the postpartum period. However, measurement of physical health is a multi-faceted approach that utilizes self-report measures and medical examination, to physical tests measuring mobility, dexterity, and pain. As a result, the first step in any investigative or research process is to confirm that the measurement is valid and free of bias.

CHAPTER 3

CURRENT STUDY

3.1 Background

Although much research has focused on depression, anxiety, and stress during the postpartum period, there is little research done on physical functioning. Physical functioning is a broad concept that refers to an individual's ability to perform various normal physical activities when in good health (Stewart, Ware, & Brook, 1978; Haley, McHorney, & Ware, 1994). As discussed by Stewart and Ware (1992) physical functioning covers three main attributes: (1) self-care, (2) mobility, and (3) other physical activities and body movements. The study of physical functioning is important because decreased function is associated with "greater physical impairment, greater disability in valued activities, and depressive symptoms" (Katula et al., 2004, pp. 19). Likewise, decreased physical functioning may lead to a sedentary lifestyle which may increase an individual's risk of other diseases (Katula et al., 2004, pp. 19). This study will not only provide more information about an already understudied topic (Webb et al., 2008), but will also offer information about other factors that could influence ethnic groups endorsement of items on the SF-12 physical functioning subscale. Moreover, precise assessment of physical functioning and overall HRQoL may improve quality of postpartum care (Webb et al., 2008) as indicated by clinicians' concern of lack of data as a significant barrier to delivery of care during the postpartum period (Kline, Martin, & Deyo, 1998).

3.2 The SF-12 and Measurement Invariance

This study examined responses to the 12-item short-form health survey (SF-12). The SF-12 is the shortened, yet valid, alternative to the SF-36. Both questionnaires are used for general and/or specific populations (Medical Outcomes Trust, 2010; RAND Health, 2010).

The 36-Item short-form health survey (SF-36) was created for the Medical Outcomes Study that assessed adherence from patients with chronic conditions. It is a self assessed HRQoL questionnaire that measures health status. Despite the SF-36 and SF-12's strong psychometric properties, it is not without its weaknesses. Fleishman and Lawrence (2003) cautioned against making comparisons of mental health among different age groups without establishing invariance across groups. In their study, these authors examined whether prior findings showing that African-Americans had better health than Caucasians when assessed with the SF-12 was due to DIF. Their results indicated the presence of DIF in items assessing mental health and that after controlling for them, differences among groups were rendered non significant. Together with findings that differences between age groups were reversed after adjusting for DIF, this strongly suggests that the invariance assumption must be tested and explained in order to obtain accurate estimates of groups. Although, the construct validity of the SF-36 and the SF-12 is sound, it may still exhibit DIF from cross-national comparisons in specific populations (Bjorner, Kreiner, Ware, Damsgaard, & Bech, 1998).

3.3 Physical Functioning and Social Support in Postpartum

A theoretically-guided analysis of the psychometric properties of the SF-12 among women at postpartum provides a unique opportunity to examine variables that can mediate the relationship between ethnicity and the endorsement of physical health items among women during the postpartum period. As stated previously, the postpartum period is marked by physical and emotional changes (Baker, Cross, Greaver, Wei, & Lewis, 2005). The transition to motherhood carries with it new roles such as caring for an infant and coping with new social changes in relationships and family structure (Leahy-Warren, 2007). In a study covering the pregnancy period for Latinas, Chasan-Taber et al. (2007) showed that older age at delivery was associated with higher levels of occupational activity, whereas, increasing parity (number of children born by one woman) was inversely related to higher levels of occupational activity. Moreover, there are various variables identified as barriers or facilitators of physical exercise

among women in postpartum. Marquez et al. (2009) found in a qualitative study that lack of energy, motivation, resources, and time were identified as barriers to exercise during pregnancy among Latina and Non-Latina women. In contrast, social support was a strong facilitator for these same women.

Social support also has been identified as a variable associated with physical functioning. Social support is support that is due to the influence of others and can be either perceived or received. As a concept it can measure structural or functional aspects of support. Structural social support is more quantitative in nature because it measures number of relationships such as marital status, number of friends, frequency of interaction with others, and number of personal roles an individual has. Whereas functional social support is more qualitative in nature because it describes the resources an individual has and measures emotional (encouragement), instrumental (material, financial, or physical), and informational support (advice, guidance) (Helgeson, 2003; Leahy-Warren, 2007; Thornton et al., 2006). In another qualitative study by Thornton et al, (2006), the role of social support in facilitating physical activity was confirmed. Informational and emotional support from husbands was amongst the most influential and consistent on postpartum Latinas physical activity practices. Conversely, social isolation from husbands proved to be a barrier to exercising.

In summary, a review of the literature on postpartum has shown specific variables, other than sociodemographic differences, may play a role in physical health among women in postpartum. Namely, social support has been shown to be an important variable associated with physical symptom burden and depressive symptoms among Caucasian, African-American, and Hispanic mothers (Howell, Mora, Horowitz, & Leventhal, 2005). Along with social support, parity and age were also found to be the most relevant variables to include in this study.

3.4 Hypotheses

In order to expand upon variables that could explain DIF among ethnic groups of women in postpartum and explain psychological variables involved in HRQoL, the following hierarchical hypotheses will be tested.

3.4.1 Hypothesis 1

DIF will be found in the SF-12 items measuring physical health across Caucasians, African-Americans, and Latinas.

3.4.2 Hypothesis 2

Social support, Age, and Parity will mediate DIF across ethnic groups.

CHAPTER 4

METHODS

4.1 Methods

4.1.1 Sample

The patient population consisted of a convenience sample of uncomplicated postpartum mothers who delivered at an urban tertiary care academic medical center between January and September 2002. Eligibility requirements included 18 years of age or older, English or Spanish speaking, delivered infants with birth weights equal to or greater than 2,500 g, and had 5-minute Apgar scores greater than 6. Exclusive criteria were hospitalization more than 3 days for vaginal deliveries and more than 5 days for cesarean delivery. Length of hospital stay was used as proxy for maternal complications because the hospital log did not reliably code complications. Initially, a nonrandom sample (*n* = 1,166) of postpartum mothers who met all eligibility requirements were approached. Seventeen percent (*n* = 202) refused to participate. Of those who consented to participate 21% were unreachable by phone (*n* = 240). Final sample of postpartum mothers who were approached and successfully completed the surveys and interview was 62% (*n* = 724). For this study only Caucasian, Latina, and African-American women were included in the analysis (*n* = 655). See Figure 4.1 for more information.

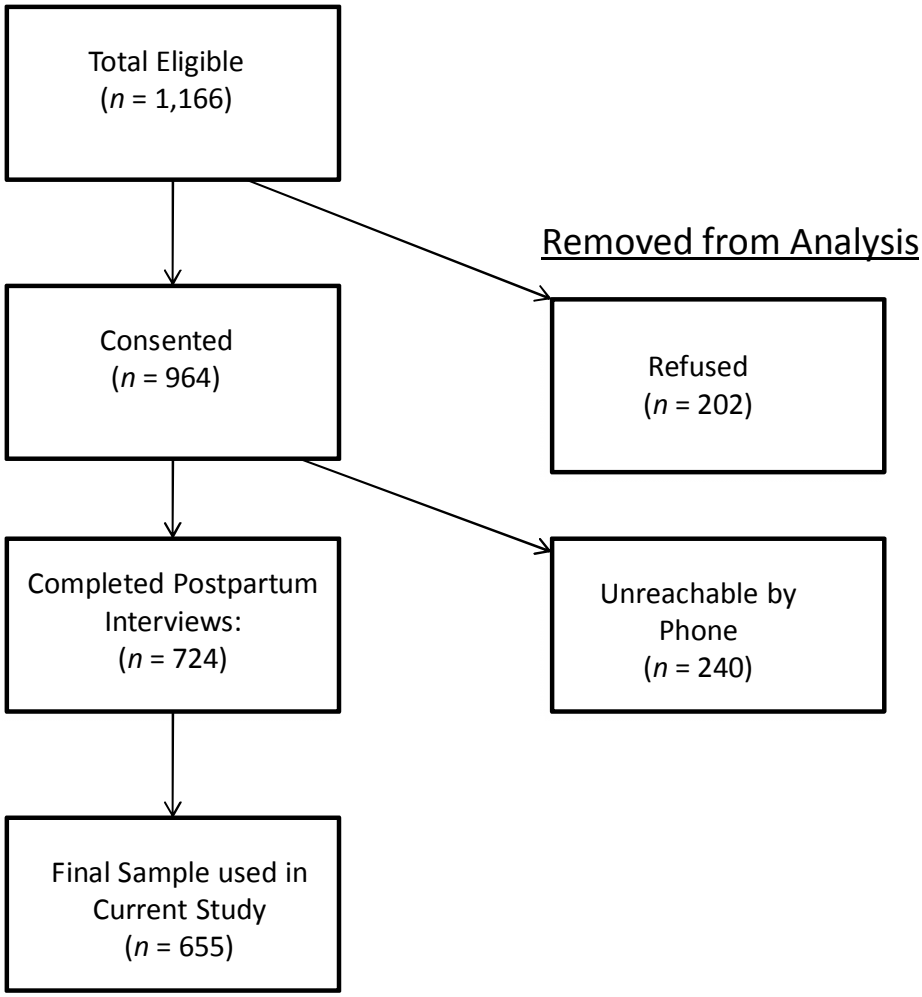

Figure 4.1 Recruitment and Follow Up in Study.

4.1.2 Design & Procedure

The data for this study was obtained from the Maternity Outcomes project (MOP: Howell, Mora, Horowitz, & Leventhal, 2005; Howell, Mora, & Leventhal, 2006; Howell, Mora, DiBonaventura, & Leventhal, 2009). The MOP was a prospective cohort longitudinal study designed to follow participants over time and examine how factors they differ on will affect certain outcome variables. Three specific data collection points were used postpartum: baseline (between 2 and 6 weeks), 3 months, and 6 months. Telephone interviews were

conducted to cover physical symptoms, emotional symptoms, daily function, social support, personal factors, perceptions of health care provider, and demographics.

After acquiring written consent from physicians, patients were identified through the Labor and Delivery log for eligibility requirements. When approached, patients signed written consent forms to participate in the study. All postpartum mothers received monetary incentives to participate. After acquiring consent, patients were contacted by telephone and interviewed for baseline between 2 and 6 weeks postpartum. All interviews lasted between 35 to 45 minutes and were conducted either in English or Spanish. Ten attempts were made to contact patients by phone; otherwise those patients who were unreachable were dropped from the study.

4.1.3 Measures

4.1.3.1 The 12-Item short-form health survey

The SF-12 (Ware, Kosinski, & Keller, 1996) is a 12-item scale that was condensed from the earlier SF-36 that measures health-related quality of life summary measures (second order factors): the physical component summary (PCS) and the mental health component summary (MCS). It consists of 12 items; item SF1 "self assessed general health"; item SF2 "moderate activities"; item SF3 "climbing stairs"; item SF4 "physical limitation"; item SF5 "physical work limitation"; item SF6 "emotional limitation"; item SF7 "emotional work limitation"; item SF8 "bodily pain"; item SF9 "calm and peaceful"; item SF10 "feel energy"; item SF11 "downheart and blue"; and item SF12 "social limitation from physical and emotional". Items SF1, SF2, SF3, SF4, SF5, and SF8 load onto PCS and items SF6, SF7, SF9, SF11, SF10, and SF12 load onto MCS. The PCS measure consists of 4 subscales (first order factors): physical functioning, role limitations caused by physical injuries, bodily pain, and general health. The MCS measure consists of 4 subscales (first order factors): vitality, social functioning, role limitations caused by emotional problems, and mental health. Participants must respond on a variety of scales, ranging from dichotomous to 5 point scales. All scores are then transformed to a standardized

0-100 scale with 50 being the mean and a standard deviation of 10. Higher scores indicate better health, function, and overall HRQoL. Construct validity for the SF-12 has been confirmed (Jenkinson, Chandola, Coulter, & Bruster, 2001). Additionally, it has been shown to be useful measures of overall community health relative to traditional measures such as total deaths or physician to population ratio (Burdine, Felix, Abel, Wiltraut, & Musselman, 2000). The psychometric properties for the SF-12 are robust. Ware, Kosinski, and Keller (1996) found correlations of .89 and .76 for the PCS and MCS, respectively, following a two week test-retest measure of reliability. Moreover, multiple R squares of .91 and .92 were found to predict SF-36 PCS and SF-36 MCS, respectively.

4.1.3.2 Social Support

This measure was developed based on results from a focus group with postpartum mothers, healthcare providers, and extensive literature searches. Pilot study with 18 women confirmed that the questions were understandable to participants. This scale consists of 4 questions that cover emotional support (in the first two weeks after your baby was born, how often did you have someone around whom you could really talk to about your feelings and who provided emotional support for you?); instrumental support with the infant (in the first two weeks, how often did you get the kind of help you needed with the baby like diapering, feeding, comforting, or holding your baby?); instrumental support with the house (in the first two weeks, how often did you get the kind of help you needed with taking care of the household, like cleaning, cooking, or grocery shopping?), and one question about partner support (looking back over the past few months how much of the time have you felt that your partner was someone you can depend on?) Participants responded on a 5-point response scale. Higher scores indicate more social support. Invariance analysis of the social support scale was conducted via MIMIC modeling (see Appendix 1). All items were free of bias and, hence, used to estimate a latent variable for social support.

4.1.3.3 Age

This variable was measured by asking participants (how old are you?) and responses were recorded on a continuous scale.

4.1.3.4 Parity

Parity is defined as the number of times a woman has given birth. Data for this variable was obtained from medical records for each participant. It was measured on a continuous scale.

4.1.3.5 Ethnicity

Ethnicity was measured by asking participants “which of the following best describes your racial or ethnic background?” and the choices included "Hispanic or Latina", "White or Caucasian", "Black or African American", "Asian", "Native American", and "Other". For the purposes of this study only Caucasian, African-American, and Hispanic participants were retained (*n*=655). This variable was then coded into two dummy variables. The first dummy code was for Hispanics and the second was for African-Americans with Caucasians being the reference group in both dummy variables.

4.1.4 Statistical Analysis

The presence of DIF was examined using a Multiple Indicator Multiple Cause (MIMIC) framework. The MIMIC latent variable model is an extension of the confirmatory factor analytic model and can be used to determine the presence of DIF (Jones & Gallo, 2002). Specifically, it is a confirmatory factor analysis *with* covariates that uses a single covariance matrix and dummy codes for group membership. The latent factor and items are regressed onto the dummy codes to indicate group differences. If there is a difference of endorsement of items then that signifies group-specific responses exist which will bias interpretation of observed means among the different groups. MIMIC models include exogenous variables that have an effect on the latent variables and on the indicator items. One advantage of MIMIC modeling is the ability to analyze multiple exogenous variables simultaneously (Fleishman, Spector, &

Altman, 2002). Thus, one can examine both the impact of the underlying latent factor on items (e.g., quality of life) and potential mediators.

The examination of DIF relies on the main assumption of unidimensionality relevant to MIMIC modeling (Cole, Kawachi, Maller, & Berkman, 2000). Unidimensionality means that the items represent a single underlying factor. Many approaches have been used to assess unidimensionality. However, in line with various researchers, such as Reise, Morizot, & Hays (2007) and Chen, West, & Sousa (2006), this assumption provides a challenge for researchers. For example, studying various sub-constructs (e.g., strength, mobility, dexterity, and speed) while at the same time trying to measure participants on one overall dimension (physical health), the unidimensionality assumption forces the researcher to fit multidimensional data to a unidimensional model, thereby, showing an inaccurate representation of the items (Reise, Morizot, & Hays, 2007). Therefore, a bifactor model is favored, especially for psychological instruments that fulfill the unidimensionality assumption while at the same time collecting multidimensional data. In the bifactor model, a general factor is used that loads unto all items (representing unidimensionality), and group factors that load only to specific items forming sub-scales (that represent multidimensionality).

Figure 4.2 represents a hypothesized model of an item showing differential item functioning (DIF) across Caucasian, African-Americans, and Hispanic women. DIF is present when the effect from the exogenous variable to an item on the scale is not influenced by the latent variables. Therefore, some other variable is influencing the endorsement of these items. The other exogenous variables in the model will be used to explain the DIF effect. These variables are social support, parity, and age.

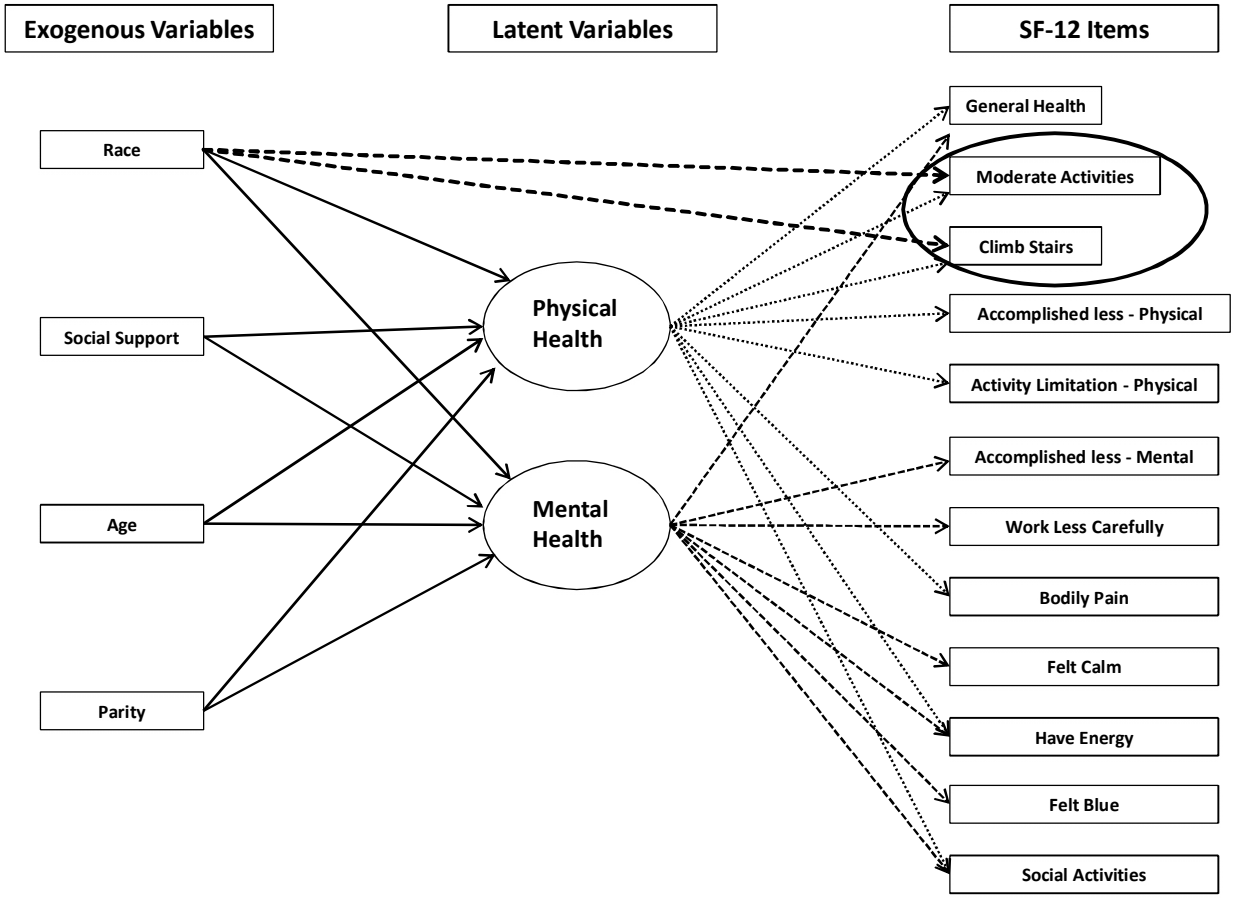

Figure 4.2 MIMIC Model of SF-12.
Observed variables are shown as rectangles while latent variables as shown as ovals. Each arrow represents a causal relationship from one construct to the other. Latent variables influence the items on the right. The two latent variables are influenced by exogenous variables on the left-hand side of the figure. However, the dashed arrow from race to "moderate activities" and "climb stairs" represents one possible DIF effect that is not mediated by the latent variables. The author used material from Fleishman and Lawrence's (2003) study for construction of this diagram in order to maintain consistency in the research field.

Parameters were estimated with Weighted Least Squares Mean and Variance adjusted (WLSMV). The WLSMV estimator assumes that missing data is missing completely at random (MCAR). Missing data modeling by WLSMV is conducted using pairwise present when there are no covariates in the model. A chi-square test was conducted in order to assess whether our data was MCAR by utilizing Little and Rubin's (2002) method. Afterward, a viable confirmatory factor analytic (CFA) model was used for MIMIC modeling. After running the CFA model, a mimic model was estimated by adding the ethnicity variables but with no direct effects. Next, paths from ethnicity variables to the hypothesized items were added but assumed to have no direct effects by constraining the pathway to zero. Then the modification indices were checked. A forward stepwise procedure was initiated by freely estimating the direct effects from the covariate to the item for the items with the highest modification index. Once no modification indices are shown, the direct effects (i.e., the DIF effects) from the covariate to the item were evaluated. After a final model was retained, the parameters were re-estimated by using the maximum likelihood estimator with robust standard errors (MLR) in order to obtain proportional odds ratio (OR) values. The OR was calculated as an effect size of the magnitude of DIF (Cole, Kawachi, Maller, & Berkman, 2000). Odds ratios with values >2.0 or <0.5 were considered a meaningful DIF effect.

Mode fit was assessed using chi-square test (χ^2), comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the weighted root mean square residual (WRMR). The following cut off points, p (χ^2) > .05, CFI and TLI > .95, RMSEA < .06, and WRMR < 1.0, will be used to determine adequate exact and close fit of models to the data (Tabachnick & Fidell, 2007; Yu, 2002). All analyses were conducted with Mplus software version 6.1.

CHAPTER 5

RESULTS

5.1 Results

5.1.1 Descriptives

Majority of the sample consisted of Caucasian females (55%) with Hispanic and African American making up approximately 29% and 16% of the sample, respectively. Median age of the participants is 31 years. Concerning parity, the majority of the sample (43%) are nulliparous with 33% as pimiparous and 24% consisting of multiparous women. Eighty one percent had a partner, and conversely, with 19% not having a partner at postpartum. Around 10% of the sample did not complete high school, with 35% of the sample completing high school, and 55% of the sample completed college or above. Sixteen percent of the sample made less than $15,000 in year, 24% made between $15,000 to $45,000, 8% made between $45,000 to $60,000, with the majority of participants (52%) making $60,000 or more. For a breakdown by race please refer to Table 5.1.

Table 5.1 Participant Descriptives (*N* = 655).

	White		Hispanic		African American	
Variables	n	%	n	%	n	%
Marital Status						
With a Partner*	356	99	128	68	47	44
Without a Partner**	5	1	59	32	60	56
Education						
Less than Junior High	0	0	7	3.7	0	0
Less than High School	4	1.1	38	20.3	14	13.1
High School or GED	30	8.3	54	28.9	33	30.8
Some College	25	6.9	54	28.9	36	33.6
College graduate/pro training	302	83.7	34	18.2	24	22.4
Annual Household Income						
Less than $5,000	3	0.8	17	9.1	11	10.3
$5,000 to $15,000	13	3.6	36	19.3	14	13.1
$15,000 to $30,000	11	3	47	25.1	31	29
$30,000 to $45,000	12	3.3	19	10.2	24	22.4
$45,000 to $60,000	20	5.5	17	9.1	8	7.5
$60,000 or more	280	77.6	27	14.4	12	11.2

*With a Partner = Married or Cohabiting
**Without a Partner = Single, Separated, Divorced, or Widowed.

5.1.2 CFA Modeling

First a missing data analysis was conducted. Little's MCAR chi-square test was used to assess whether the data was missing completely at random (MCAR). Results revealed that the data for the SF-12 was missing completely at random, χ^2 MCAR (*df* = 13694388) = 98.95, *p* = 1.0. Therefore all data was used for subsequent analysis. Item descriptives of the SF-12 are shown in Table 5.2.

Table 5.2 Item Descriptives of SF-12

Short Form Health *Survey*	Response Frequencies (With Percentages)	Mean	SD
SF1 - General Health		2.52	1.16
Excellent	161 (24.6)		
Very Good	156 (23.8)		
Good	211 (32.2)		
Fair	91 (13.9)		
Poor	35 (5.3)		
SF2 - Moderate Activities		1.96	0.85
Limit you a lot	248 (37.9)		
Limit you a little	180 (27.5)		
Not limit you at all	225 (34.4)		
SF3 - Climbing Stairs		2.12	0.87
Limit you a lot	198 (30.2)		
Limit you a little	146 (22.3)		
Not limit you at all	272 (41.5)		
SF4 - Physical Limitation		1.46	0.50
Yes	355 (54.2)		
No	297 (45.3)		
SF5 - Physical Work Limitation		1.38	0.49
Yes	403 (61.5)		
No	250 (38.2)		
SF6 - Emotional Limitation		1.82	0.39
Yes	123 (18.8)		
No	531 (81.1)		
SF7 - Emotional Work Limitation		1.87	0.34
Yes	84 (12.8)		
No	567 (86.6)		
SF8 - Bodily Pain		2.44	1.33
Extremely	64 (9.8)		
Quite a bit	103 (15.7)		
Moderately	98 (15.0)		
A little bit	183 (27.9)		
Not at all	206 (31.5)		

Table 5.2 – *Continued*

SF9 - Calm and Peaceful		3.40	1.45
All of the time	59 (9.0)		
Most of the time	183 (27.9)		
A good bit of the time	56 (8.5)		
Some of the time	192 (29.3)		
A little of the time	121 (18.5)		
None of the time	43 (6.6)		
SF10 - Feel Energy		4.32	1.40
All of the time	28 (4.3)		
Most of the time	72 (11.0)		
A good bit of the time	47 (7.2)		
Some of the time	163 (24.9)		
A little of the time	203 (31.0)		
None of the time	141 (21.5)		
SF11 - Downheart and Blue		4.92	1.14
All of the time	12 (1.8)		
Most of the time	25 (3.8)		
A good bit of the time	16 (2.4)		
Some of the time	140 (21.4)		
A little of the time	222 (33.9)		
None of the time	240 (36.6)		
SF12 - Social Limitation From Physical and Emotional		4.03	1.26
All of the time	39 (6.0)		
Most of the time	64 (9.8)		
Some of the time	84 (12.8)		
A little of the time	121 (18.5)		
None of the time	346 (52.8)		
N	655		

Model building was initiated using two different factor structures of the SF-12. The original factor structure (Ware, Kosinski, Keller, 1996) and a newer factor structure based on Fleishman & Lawrence (2003) study. The original factor structure included factors Physical

Health Component Summary (PCS) and Mental Health Component Summary (MCS) with items SF1, SF2, SF3, SF4, SF5, and SF8 loading onto PCS and items SF6, SF7, SF9, SF11, SF10, and SF12 loading onto MCS. The newer factor structure had items SF1, SF2, SF3, SF4, SF5, SF8, SF10, and SF12 loading onto PCS. For the MCS factor the following items were SF1, SF6, SF7, SF9, SF10, SF11, and SF12. Several models were conducted; a standard unidimensional model (Model 1), a two factor model using the original factor structure (Model 2), a two factor model using the new factor structure (Model 3), a bifactor model using the original factor structure (Model 4), and lastly a bifactor model using the new factor structure (Model 5).

Models 1, 2, and 3 were nested within models 4 and 5. In all cases the bifactor models showed better fit via significance of the chi-square difference test. For example, the bifactor model with the original factor structure (Model 4) showed better fit to the data than the unidimensional model (Model 1), the original two factor model (Model 2), and the new two factor model (Model 3) ($\chi^2_{diff}(12) = 398.75$, $p < .000$; $\chi^2_{diff}(11) = 146.27$, $p < .000$; $\chi^2_{diff}(8) = 39.52$, $p < .000$; respectively). The bifactor model with the new factor structure (Model 5) also showed statistically significant better fit than the unidimensional model (Model 1), the original two factor model (Model 2), and the new two factor model (Model 3) ($\chi^2_{diff}(15) = 441.94$, $p < .000$; $\chi^2_{diff}(14) = 181.12$, $p < .000$; $\chi^2_{diff}(11) = 75.06$, $p < .000$; respectively). For this reason, the bifactor models were the most useful in providing the best fit for the data. When comparing the two bifactor models, the bifactor model with the new factor structure (Model 5) showed better fit compared to the original factor structure ($\chi^2_{diff}(3) = 39.60$, $p < .000$). Model 5 was used as the base model for MIMIC analyses (Figure 5.1). See table 5.3 for more information.

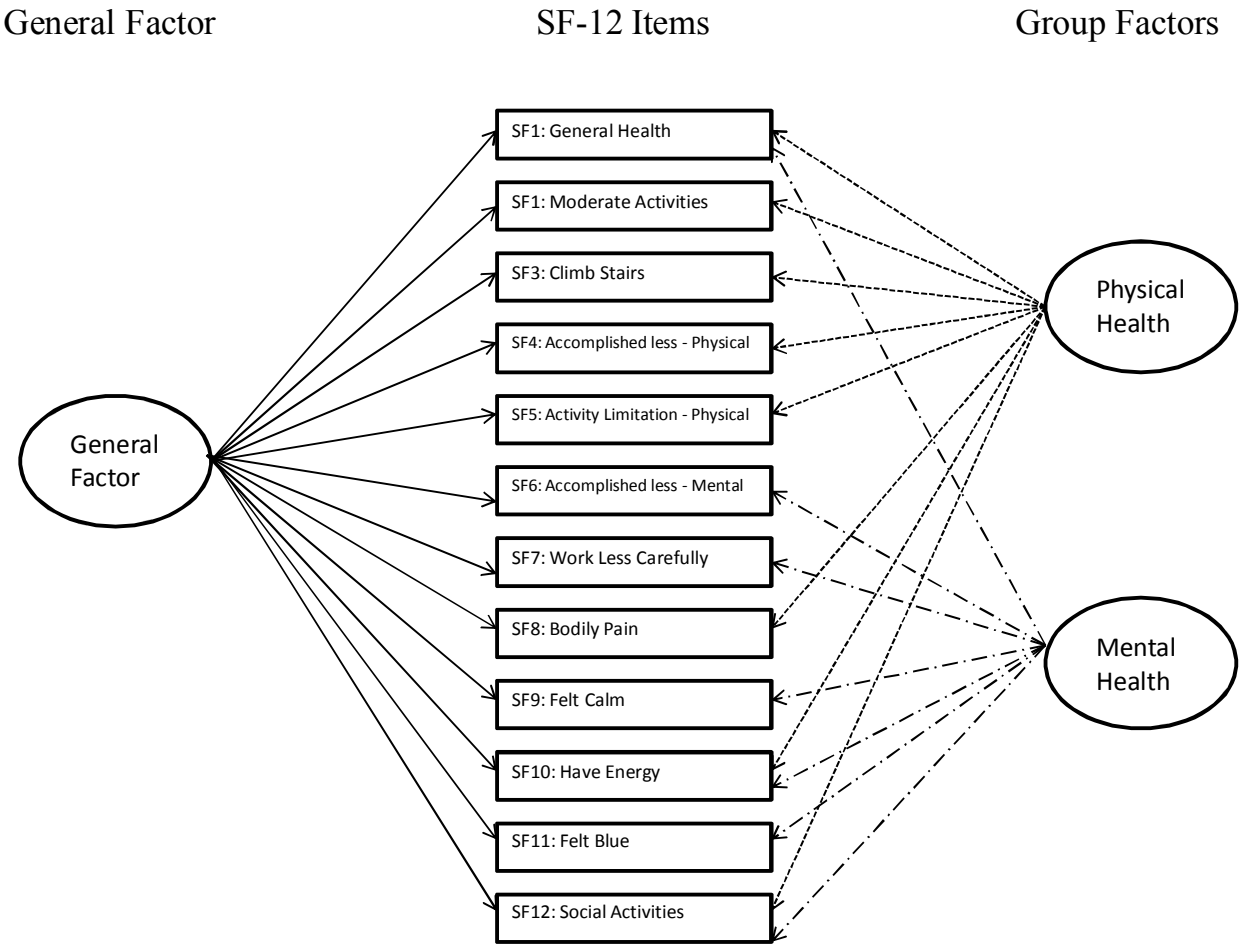

Figure 5.1 Bifactor Model with New Factor Structure.

Table 5.3 SF-12 CFA Models

	χ^2	*df*	*p-value*	RMSEA	WRMR	CFI	TLI
Model 1: One Factor (All Categorical)	783.46	54	0.0000	.144	2.579	.852	.820
Model 2: Two Factors (All Categorical) *Original* Factor Structure	373.31	53	0.0000	.096	1.735	.934	.919
Model 3: Two Factors (All Categorical) *New* Factor Structure	219.68	50	0.0000	.072	1.237	.966	.955
Model 4: Bifactor Model *Original* Factor Structure	189.13	42	0.0000	.073	1.044	.970	.953
Model 5: Bifactor Model *New* Factor Structure	141.56	39	0.0000	.063	.882	.979	.965

5.1.3 MIMIC Model

For the first MIMIC model all items were regressed on the dummy coded variables that denoted group membership of Hispanics and African-Americans using a Bifactor model, χ^2 (57) = 254.12, p = 0.000, RMSEA = .073, WRMR = 1.10, CFI = .961, and TLI = .939. The modification indices specified that the model would be improved by freely estimating direct effects of ethnicity on the following items: SF9 “Calm and Peaceful”, (2) SF8 “Bodily pain”, and (5) SF1 “General Health”. The estimation was done in a stepwise manner, that is, the parameters were freed for one item at the time.

First the impact of ethnicity on SF9 was freely estimated. The regression from Race denoting Hispanic membership to SF9 revealed a statistically significant estimate of 0.47 and a standard error (SE) of 0.095. The impact of African-American group membership on SF9 was also statistically significant, B = 0.36, SE = 0.103.

In the next step, the direct effect of Hispanics and African-American ethnicities on SF8 “Bodily pain” was then freely estimated. The impact of both Hispanic (B = 0.25, SE = 0.079) and African-American (B = 0.28, SE = 0.094) ethnicities were statistically significant on the item response of SF8 “Bodily pain”.

The last MIMIC model freely estimated the direct effects of both ethnic groups on items SF9, SF8, and SF1. The DIF effects from both ethnic groups to all three items were significant. The previous estimates changed a little in this model but the differences were considered inconsequential. The Hispanic and African-American group membership on SF1 is B = -0.27, SE = 0.101; B = -0.27, SE = 0.114, respectively.

After MIMIC direct effects were identified a model using the MLR estimator was conducted in order to extract effect sizes for DIF. The MLR estimator was used in conjunction with an algorithm integration using the Monte Carlo method and convergence set at 0.000001 and a log criterion set at 0.000001.

The DIF effects from the covariate of Hispanic ethnicity to SF1 was non-meaningful (OR = 0.55, CI_{95} = 0.36, 0.84) and also had a nominal effect size for African-Americans (OR = 0.56, CI_{95} = 0.32, 1.01). The DIF effect from Hispanics to SF8 had a small meaningful magnitude (OR = 1.8, CI_{95} = 1.19, 2.72). For African-Americans the DIF effect was significant and had a meaningful effect (OR = 2.11, CI_{95} = 1.20, 3.71). Finally, the DIF effects from Hispanics to SF9 had a meaningful effect size (OR = 2.62, CI_{95} = 1.64, 4.17). On the other hand, the group effect of African-Americans unto SF9 was non-meaningful (OR = 1.41, CI_{95} = 0.68, 2.95).

This evidence revealed that from the six DIF effects identified, only African-Americans' differential endorsement of item SF8 and Hispanics endorsement of item SF9 were meaningful and relevant DIF effects (See Figure 5.1 and 5.2 for Item Characteristic Curves). However, in accordance with Yang, Tommet, & Jones (2009) total effects were calculated to explain how much of the variance is accounted for by the latent factors vs. the covariates for each item exhibiting DIF ($\frac{DIF\ Effect}{Total\ Effect}$). This was done to account for some inherent limitations in using one method to examine DIF relationships.

Utilizing the standardized parameter estimates it was found that the DIF effect from Hispanic group membership to SF1 is 47% of the total effect, while African-American group membership to SF1 explained 21% of the variance of the total effect. Concerning the pain item SF8, Hispanic ethnicity explained 32% while African-American ethnicity accounted for 50% of the variance in the total effect. Finally, the DIF effect from Hispanic ethnicity to SF9 is 73% and 39% for African-Americans.

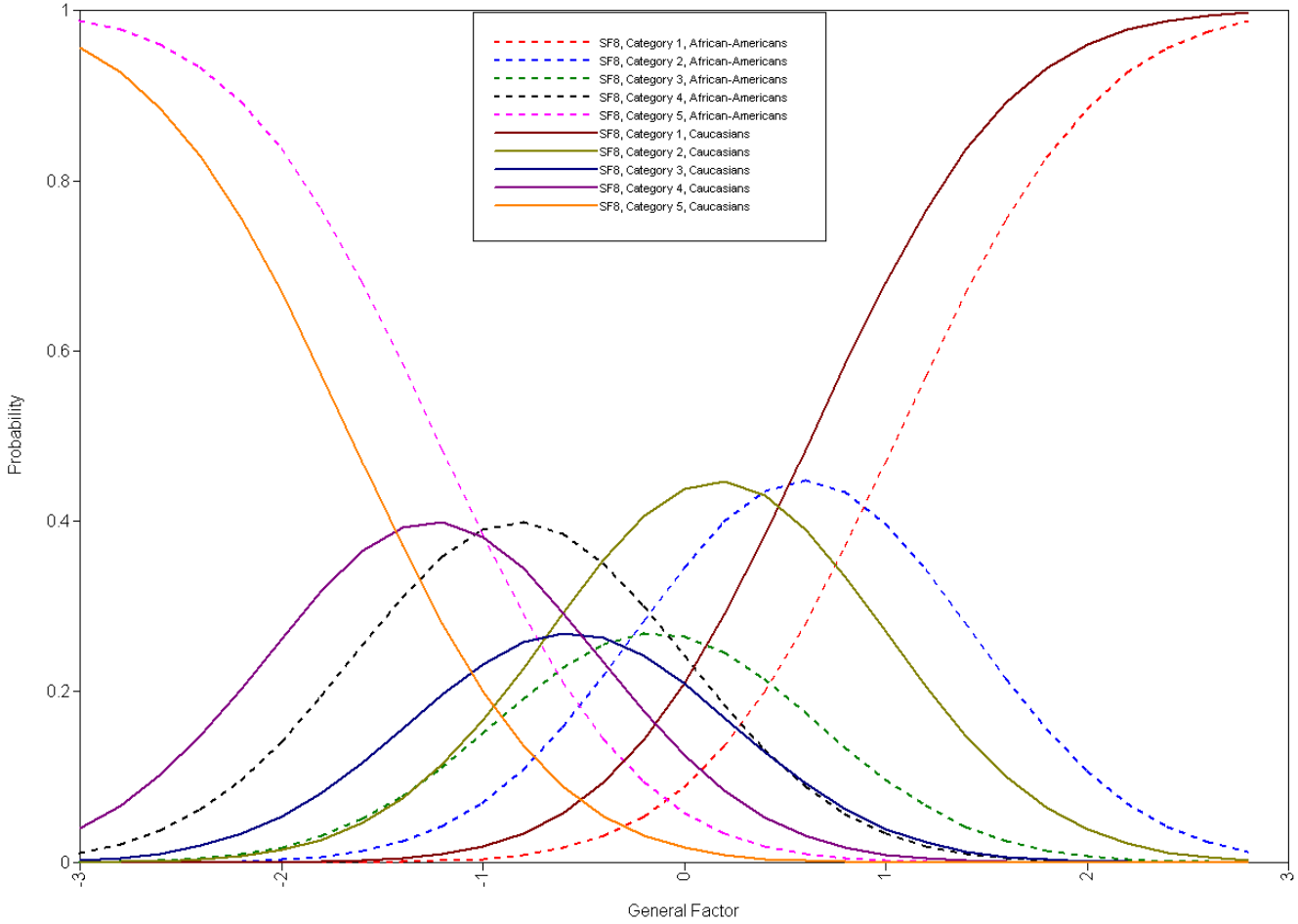

Figure 5.2 Item Characteristic Curve (ICC) of Item SF8 “Bodily Pain”
This graph represents the probabilistic response characteristics for each of the 5 categories in item SF8 as a function of the general factor (functional status) for both African-Americans and Caucasians. Data was plotted from the MIMIC Model using the WLSMV estimator.

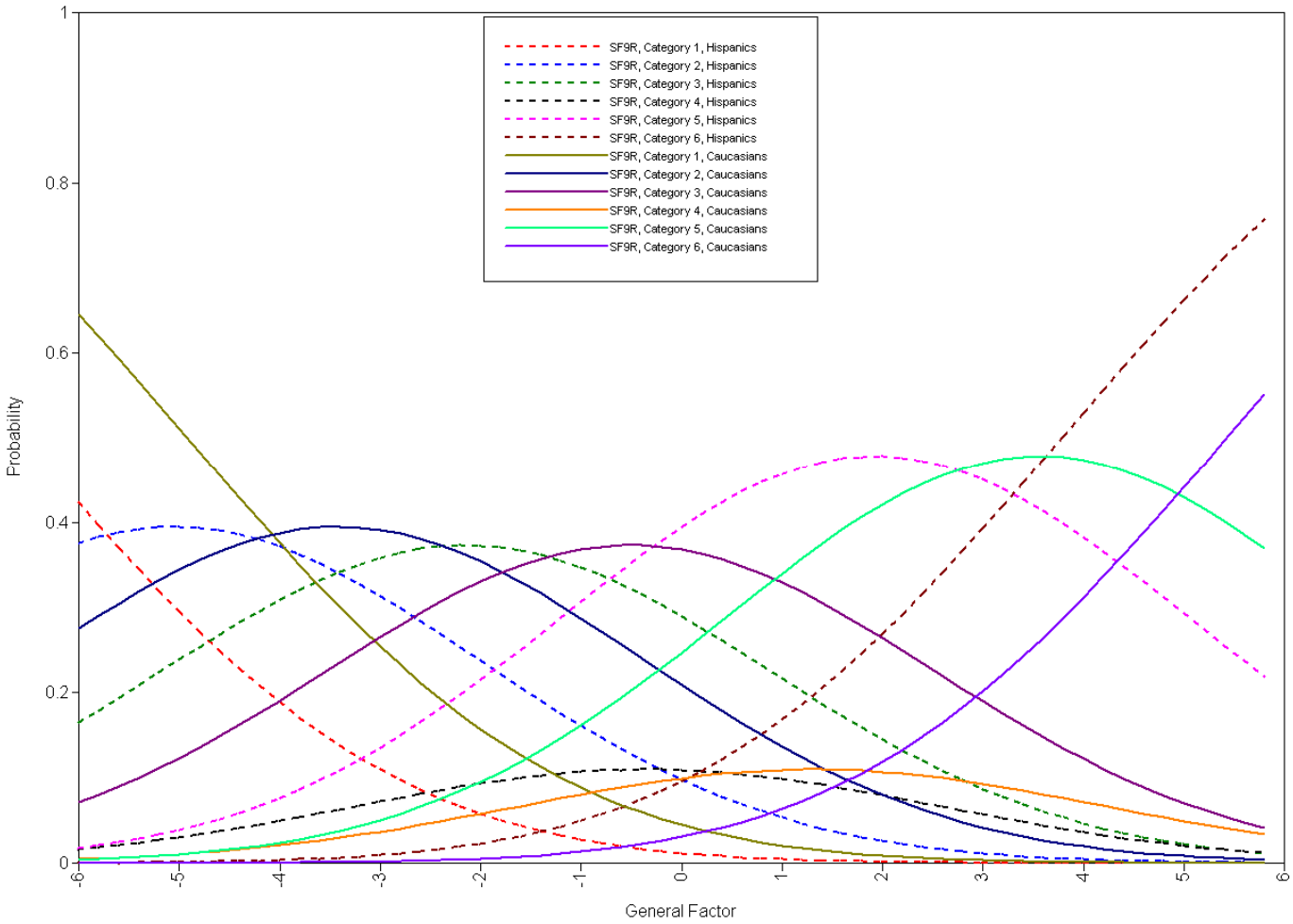

Figure 5.3 Item Characteristic Curve (ICC) of Item SF9 “Calm and Peaceful”
This graph represents the probabilistic response characteristics for each of the 6 categories in item SF9 as a function of the general factor (functional status) for both Hispanics and Caucasians. Data was plotted from the MIMIC Model using the WLSMV estimator.

5.1.4 Mediation Analysis

A mediation analysis was used to explore if other postpartum variables explained the DIF effects from Hispanics to item SF1 “self assessed general health” and African-Americans to item SF8 “bodily pain”. So as to fully explore and understand the DIF effects three theoretical relevant variables were included in a mediation model in order to mediate the DIF effects. The following three mediators were used: (1) Social Support, (2) Age, and (3) Parity.

The mediation analysis was conducted according to steps outlined by MacKinnon (2008). The independent variable (X), the mediator (M), and the dependent variable (Y) are included in a model with the following pathways estimated (c, a, b, c'). The c pathway is the total effect of X on dependent variable Y. The a pathway represents the effect of X on the mediator M. The b pathway is the effect of mediator M on outcome variable Y. The c' pathway represents the effect of X on Y when controlling for the mediator M. The ab estimate, which is the product of pathways a and b, is the indirect effect, also known as the mediation effect.

5.1.4.1 Item 1 “Self Assessed General Health” Single Mediation Model

Originally, a mediation analysis was conducted with three mediators: (1) Social Support, (2) Age, and (3) Parity. Age and Parity were dropped out from the analysis because they did not relate to the item. Consequently, only social support was used as a mediator. Ethnicity had an effect on Social support (Hispanic a = -0.53, p < .001; African-American a = -0.62, p < .001). The social support mediator had an effect on item SF1 (b = 0.17, p < .01). Taken together, the analysis shows that social support partially mediated the DIF effect of Ethnicity on SF1. Effect size for the mediation effect ($\frac{ab}{c}$) was computed. The mediation effect of social support accounted for 48% of the total effect for Hispanic ethnicity on item SF1 and accounted for 77% of the total effect for African-American ethnicity on item SF1. See Table 4 in the Appendix for more details.

5.1.4.2 Item 8 “Bodily Pain” Multiple Mediation Model

A multiple mediation analysis was conducted with two mediators: (1) Social Support, and (2) Parity. The mediating variable of “Age” was dropped from the model after finding a non-significant relationship close to zero. Social support mediated the effect of Hispanic ethnicity on item SF8 (a = -0.56, p < .001; b = -0.13, p < .01) and for African-American ethnicity on SF8 (a = -0.63, p < .001; b = -0.13, p < .01). The mediation effect of social support accounted for 38% of the total effect for Hispanic ethnicity and SF8. For African-Americans the mediation effect explained 37% of the variance in the total effect. Parity did not mediate the effect of Hispanic ethnicity on item SF8 (a = 0.17, ns; b = -0.11, p < .001). However, for the effect between African-American ethnicity and the SF8 item, parity did mediate the effect (a = 0.45, p < .001; b = -0.11, p < .001). Though, the mediation effect of parity was not significant for the effect of Hispanic ethnicity on SF8 item, it did account for 23% of the total effect of African-American ethnicity on the SF8 item. See Table 5 in the Appendix for more information.

CHAPTER 6

DISCUSSION

The first aim of this research study was to determine measurement invariance of the item measuring physical health based on important relevance to HRQoL. Three other items did reveal noninvariance. Subsequently, the items that were related to the general hypothesized relationships were further explored using a mediation analysis to explain the subgroup variation found.

First, several models were conducted to assess the best fit of the data. By and large, the bifactor model yielded the best fit for the SF-12 data. Several authors have listed key advantages of the bifactor model over second-order models and multi-dimensional models. The notable advantage being the ease of DIF interpretations due to the bifactor model separating the unique contributions of the relationships from either a general dimension or the second order factors (Reise, Morizot, & Hays, 2007; Chen, West, & Sousa, 2006).

The three items that did exhibit DIF were items SF1 “self assessed general health”, SF8 “bodily pain”, and SF9 “calm and peaceful”. Specifically, both Hispanics and African-Americans were less likely to endorse the “self assessed general health” item compared to Caucasians. However, the effect sizes were not meaningful for the SF1 DIF effect. Nevertheless, in order to fulfill the original aims of this study a mediation analysis was conducted using social support to account for the relationship between the ethnicity covariate and the SF1 item. Social support partially mediated the DIF for both Hispanics and African-Americans. Specifically, the more social support both ethnic groups received, the more likely they are to rate their overall general health in a positive manner ($\beta = 0.17$, $p < .05$). This finding is in agreement with research discussed previously (Thornton et al., 2006; Howell, Mora, Horowitz, & Leventhal, 2005; Marquez et al., 2009).

For item SF8 “bodily pain”, Hispanic group membership endorsement of the pain interference item was found to be non-relevant upon further examination of the effect size. Conversely, results showed that African-Americans were more likely to endorse this item more than Caucasians. Moreover, this relationship was meaningful based on the effect size and explained 50% of the variance in the total effect. Using the two mediators of social support and parity the results confirmed a partial mediation. Specifically, the more social support both ethnicities received the less likely they are to report pain interfering with their work activities (β = -0.13, p < .05). Additionally, the more they gave birth previously the less likely they endorsed that pain interfered with their work activities in postpartum (β = -0.11, p < .05). This finding is consistent with sensation theory and more recently the common sense model (Johnson, 1973; Leventhal et al., 2004). Individuals that have had experience with a physically painful, discomforting, or stressful medical procedure or event and can anticipate the sequence of actions and sensations that will be felt are more likely to report less distress, need fewer medications, have stable cardiovascular functioning, and overall better coping mechanisms than those without prior experience or those who did not receive educational instruction (Johnson, Leventhal, & Dabbs, 1971; Johnson, Morrissey, & Leventhal, 1973; Johnson & Leventhal, 1974).

There are several limitations of this study to take into account. First, MIMIC modeling assumes that other parameters, such as factor loadings and factor variances/covariances, are invariant across the groups. Testing these other sources of invariance requires a large sample size in a Multi-Group CFA framework. Because our sample size was small (*n*=655), this limited the research to only conducting strong factorial (scalar) test of indicator intercept invariance (Brown, 2006). However, MIMIC provides a key advantage over other models by accommodating to small sample sizes because of its use of a single covariance matrix. Second, MIMIC does not detect non-uniform DIF which limits this study to only detecting uniform DIF. Third, the sample composition consisted of a majority of women with a partner (81%) compared

to 19% without a partner. This may affect the generalizability of our findings especially in the context of one item measuring social support (Partner Support) affecting response performance on certain items of the SF-12. Fourth, because this field is constantly developing, parameter estimation, effect sizes, fit indices, software applications, and other means of evaluation DIF has not been completely standardized in the field of Measurement Invariance.

Relevance of this study is as follows. First, two items did show significant and meaningful bias. After matching ethnic groups on physical health, African-American mothers were more likely to endorse the SF8 “bodily pain” item. Also, Hispanic mothers were more likely to endorse the SF9 “calm and peaceful” item after being matched on mental health compared to Caucasians. Practical implications show that if accounting for other variables, such as social support in measurement invariance analysis, subsequent findings may be explained without the call for alarm that such items are biased. Detection of bias, without understanding the effect, can call for unnecessary deletion or modification of items. By conducting a mediation analysis one will find other variables accounting for a portion of the DIF effect and thereby explaining a portion of the discrepancy observed in responses on some items. With this study’s approach, taking into account these variables may provide more information on how to administer these tests to different groups.

In conclusion, a multipronged approach using significance tests, effect sizes, unique variance estimation of the total effect, and mediation analysis may show additional variables, other than the latent factors, as partially influencing responses on certain items of the SF-12 if there is DIF. However, almost all the investigative methods did not show a strong enough effect to warrant a strong claim for DIF except for African-Americans differentially endorsing item 8 “bodily pain” and Hispanics differentially endorsing item 9 “calm and peaceful”. The results of this study show the SF-12 as a reliable measure of Physical and Mental Health for women in postpartum only after identifying important postpartum relevant variables as affecting endorsement of the items.

APPENDIX A

SOCIAL SUPPORT: DESCRIPTIVES, MODELS, AND MEDIATION ANALYSIS

Table A.1. Item Descriptives of Social Support

	Response Frequencies (With Percentages)					Mean	SD
	All of the Time	Most of the Time	Sometimes	Rarely	Never		
1. Infant Instrumental Support	289 (44.1)	151 (23.1)	143 (21.8)	28 (4.3)	43 (6.6)	2.06	1.19
2. Household Instrumental Support	354 (54.0)	140 (21.4)	101 (15.4)	14 (2.1)	45 (6.9)	1.86	1.18
3. Emotional Support	360 (55.0)	142 (21.7)	90 (13.7)	32 (4.9)	29 (4.4)	1.82	1.12
4. Partner Support	358 (54.7)	203 (31.0)	45 (6.9)	18 (2.7)	19 (2.9)	1.66	0.94
N	655						
α Coefficient	.57						
χ^2 MCAR	(df = 2620) = 152.36, *p* = 1.0						

Table A.2. Social Support CFA Models

	χ^2	*df*	*p-value*	RMSEA	WRMR	CFI	TLI
1 Factor Model	Not Indentified	NA	NA	NA	NA	NA	NA
Bifactor Model	Not Indentified	NA	NA	NA	NA	NA	NA
Rasch Model (Factor Loadings Constrained to be equal)	19.73	5	0.0014	.067	.845	.952	.952
Unconstrained Factor Loadings Model	19.22	2	0.0001	.115	.771	.944	.861
MIMIC Model for Ethnic Invariance	22.21	8	0.0045	.052	.813	.954	.936

Table A.3. Social Support MIMIC Models

Model	Estimator	DIF Effect	Value
MIMIC	WSLMV	Hispanic → Infant Instrumental Social Support	B = -0.30, SE = 0.10
MIMIC	MLR	Hispanic → Infant Instrumental Social Support	OR = 0.57, CI_{95} = 0.38, 0.85

Table A.4. Single Mediation Model of SF1 Item

	c	*a*	*b*	*c'*	*ab*
Model 1	-0.19	-0.53***	0.17**	-0.10	-0.09*
Model 2	-0.07	-0.62***	0.17**	0.03	-0.10*

Note: Unstandardized estimates are used in this table.
Model 1: X = Hispanic, M = Social Support, Y = Item SF1
Model 2: X = African-American, M = Social Support, Y = Item SF1
a = effect of X on M, b = effect of M on Y, c = total effect of X on Y, c' = effect of X on Y controlling for M, ab = indirect effect
*p < .05; **p < .01; ***p < .001

Table A.5. Multiple Mediation Model of SF8 Item

	c	a_1	b_1	*c'*	ab_1	a_2	b_2	ab_2
Model 1	0.19*	-0.56***	-0.13**	0.12	0.07*	0.17	-0.11***	-0.02
Model 2	0.21*	-0.63***	-0.13**	0.13	0.08*	0.45**	-0.11***	-0.05*

Note: Unstandardized estimates are used in this table.
Model 1: X = Hispanic, M_1 = Social Support, M_2 = Parity, Y = Item SF8
Model 2: X = African-American, M_1 = Social Support, M_2 = Parity, Y = Item SF8
a = effect of X on M, b = effect of M on Y, c = total effect of X on Y, c' = effect of X on Y controlling for M, ab = indirect effect
*p < .05; **p < .01; ***p < .001

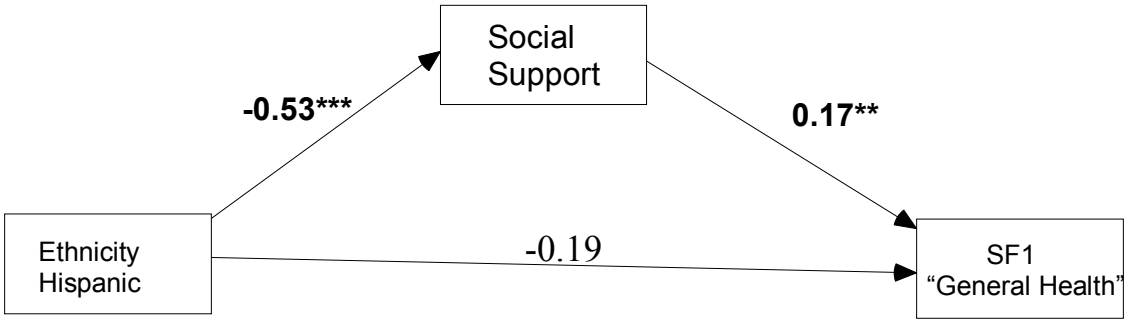

Figure A.1. Partial Mediation of the DIF effect (Hispanic → SF1) using social support as a mediator.

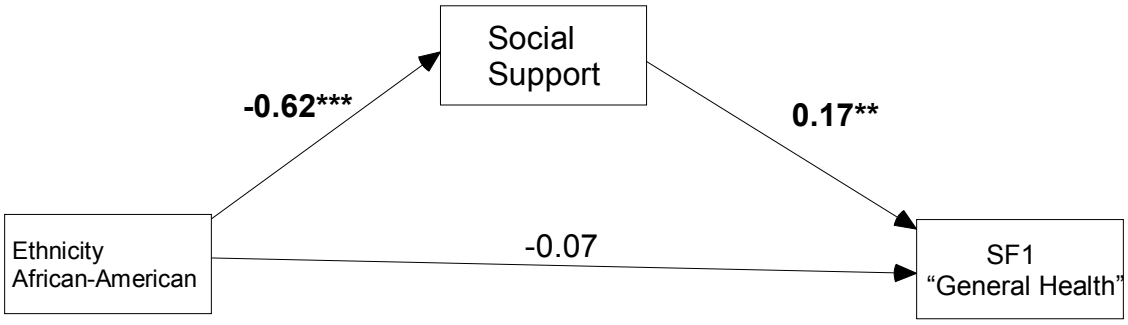

Figure A.2. Partial Mediation of the DIF effect (African-American → SF1) using social support as a mediator.

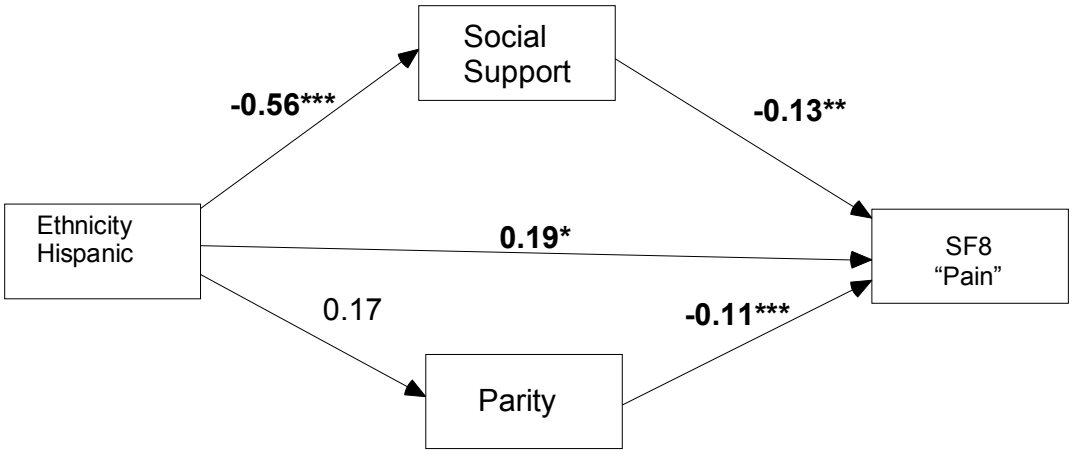

Figure A.3. Partial Mediation of the DIF effect (Hispanic → SF8) using social support as a mediator.

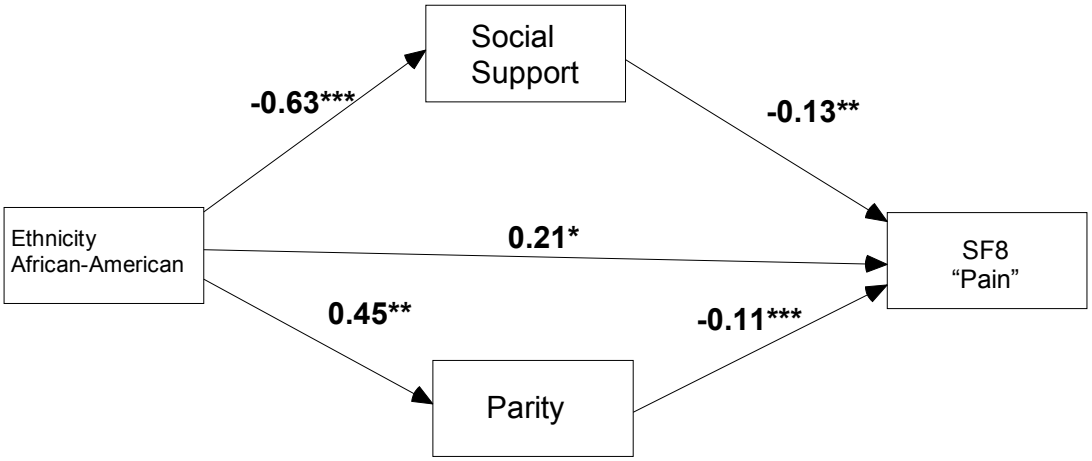

Figure A.4. Partial Mediation of the DIF effect (African-American → SF8) using social support and parity as mediators.

APPENDIX B

QUESTIONNAIRES

The 12-ITEM SHORT-FORM HEALTH SURVEY

Now we would like to ask you some questions about how you were feeling and your experiences in the first two weeks after your baby(s) was born.

1. In the first two weeks after your baby(s) was born, in general would you say your health was: (CHOOSE ONE ONLY)

Poor	Fair	Good	Very good	Excellent
○	○	○	○	○

Now we are going to ask you some questions about your health and how you were feeling those first two weeks after your baby(s) was born. The following items are about activities you might do during a typical day.

Please tell me if your health limited you a lot, limited you a little, or did not limit you at all in these activities in the first two weeks after your baby(s) was born.

2. Moderate activities, such as moving a table, pushing a vacuum cleaner, carrying groceries, walking 10 blocks. Did your health: (CHOOSE ONE ONLY)

Limit you a lot	Limit you a little	Not limit you at all
○	○	○

3. Climbing several flights of stairs. Did your health: (CHOOSE ONE ONLY)

Limit you a lot	Limit you a little	Not limit you at all
○	○	○

The following two questions ask you about your physical health and your daily activities during the first two weeks after your baby(s) was born.

4. Did you accomplish less than you would like as a result of your physical health? (CHOOSE ONE ONLY)

Yes	No
○	○

5. Were you limited in the kind of work or other regular daily activities you did as a result of your physical health? (CHOOSE ONE ONLY)

Yes	No
○	○

The following two questions ask about your emotions and your daily activities during the first two weeks after your baby(s) was born.

6. Did you accomplish less than you would like as a result of any emotional problems, such as feeling depressed or anxious? (CHOOSE ONE ONLY)

Yes	No
○	○

7. Did you not do work or other regular activities as carefully as usual as a result of any emotional problems, such as feeling depressed or anxious? (CHOOSE ONE ONLY)

Yes	No
○	○

---

8. In the first two weeks after your baby(s) was born, how much did pain interfere with your normal work, including both work outside the home and inside of the home? Did it interfere: (CHOOSE ONE ONLY)

Extremely	Quite a bit	Moderately	A little bit	Not at all
○	○	○	○	○

---

9. How much of the time in those first two weeks did you feel calm and peaceful? (READ CATEGORIES ONLY IF NECESSARY) (CHOOSE ONE ONLY)

None of the time	A little of the time	Some of the time	A good bit of the time	Most of the time	All of the time
○	○	○	○	○	○

---

10. How much of the time in those first two weeks did you have a lot of energy? (CHOOSE ONE ONLY)

None of the time	A little of the time	Some of the time	A good bit of the time	Most of the time	All of the time
○	○	○	○	○	○

---

11. How much of the time did you feel downhearted and blue? (CHOOSE ONE ONLY)

All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time
○	○	○	○	○	○

---

12. During those first two weeks, how much of the time did your physical health or emotional problems interfere with your social activities like visiting with friends, relatives, etc.? Did it interfere: (CHOOSE ONE ONLY)

All of the time	Most of the time	Some of the time	A little of the time	None of the time
○	○	○	○	○

SOCIAL SUPPORT QUESTIONNAIRE

INFANT INSTRUMENTAL SUPPORT

In the first two weeks, how often did you get the kind of help you needed with the baby(s) like diapering, feeding, comforting, or holding your baby(s)?

Never	Rarely	Sometimes	Most of the time	All of the time
○	○	○	○	○

HOUSEHOLD INSTRUMENTAL SUPPORT

In the first two weeks, how often did you get the kind of help you needed with taking care of the household (like cleaning, cooking, or grocery shopping)?

Never	Rarely	Sometimes	Most of the time	All of the time
○	○	○	○	○

EMOTIONAL SUPPORT

In the first two weeks after your baby(s) was born, how often did you have someone around whom you could really talk to about your feelings and who provided emotional support for you?

Never	Rarely	Sometimes	Most of the time	All of the time
○	○	○	○	○

PARTNER SUPPORT

Looking back over the past few months how much of the time have you felt that your partner was someone you can depend on?

Not applicable because I do not have a partner	Never	Rarely	Sometimes	Most of the time	All of the time
○	○	○	○	○	○

REFERENCES

Alexander, J., Levy, V., & Roch, S. (1990). Emotional problems associated with childbirth. *Postnatal care: a research-based approach;* (p. 45). London: Macmillan.

Baker, L., Cross, S., Greaver, L., Wei, G., & Lewis, R. (2005). Prevalence of Postpartum Depression in a Native American Population. *Maternal and Child Health Journal, 7(1),* 21-25. doi: 10.1007/s10995-005-2448-2.

Belsky, J. & Rovine, M. (1990). Patterns of Marital Change across the Transition to Parenthood: Pregnancy to Three Years Postpartum. *Journal of Marriage and Family*, *52*(1), 5-19.

Bjorner, J.B., Kreiner, S., Ware, J.E., Damsgaard, & M.T., Bech, P. (1998). Differential item functioning in the Danish translation of the SF-36. *Journal of Clinical Epidemiology, 51,* 1189–1202.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Burdine, J.N., Felix, M.R., Abel, A.L., Wiltraut, C.J., & Musselman, Y.J. (2000). The SF-12 as a population health measure: an exploratory examination of potential for application. *Health Services Research, 35(4),* 885-904.

Chasan-Taber, L., Schmidt, M.D., Pekow, P., Sternfeld, B., Manson, J., & Markenson, G. (2007). Correlates of physical activity in pregnancy among Latina women. *Maternal and Child Health Journal, 11,* 353–363.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A Comparison of Bifactor and Second-Order Models of Quality of Life. Multivariate Behavioral Research, 41(2), 189 — 225. DOI: 10.1207/s15327906mbr4102_5.

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(3)*, 285-289.

Cook, B.L., McGuire T, & Miranda J. (2007). Measuring trends in mental health care disparities, 2000-2004. *Psychiatric Services, 58(12)*, 1533-1540.

Dobalian, A., & Rivers, P. A. (2008). Racial and Ethnic Disparities in the Use of Mental Health Services. *The Journal of Behavioral Health Services and Research, 35(2)*. doi 10.1007/s11414-007-9097-8.

Finch, W. H., & French, B. F. (2008). Anomalous Type I Error Rates for Identifying One Type of Differential Item Functioning in the Presence of the Other. *Educational and Psychological Measurement, 68(5)*, 742-759. doi: 10.1177/0013164407313370.

Fleishman, J. A., & Lawrence, W. F. (2003). Demographic Variation in the SF-12 Scores: True Differences or Differential Item Functioning. *Medical Care, 41(7)*, S75-S86.

Fleishman, J.A., Spector, W.D., & Altman, B.M. (2002). Impact of differential item functioning on age and gender differences in functional disability. *Journal of Gerontology: Social Sciences, 57B(5)*, S275-S284.

Gallo, J. J., Anthony, J. C., & Muthén, B. O. (1994). Age differences in the symptoms of depression: a latent trait analysis. *Journals of Gerontology, Psychological Sciences, 49*, 251-264.

Gallo, J. J., & Rabins, P. V. (1999). Depression without sadness: alternative presentations of depression in late life. *American Family Physician, 60*, 820-860.

Haley, S. M., McHorney, C. A., & Ware, J. E. (1994). Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): I. Unidimensionality and Reproducibility of the Rasch Item Scale. *Journal of Clinical Epidemiology, 47(6)*, 671-684.

Helgeson, V. S. (2003). Social support and quality of life. *Quality of Life Research, 12(1)*, 25-31.

Howell, E., Mora, P. A., DiBonaventura, M. D., & Leventhal, H. (2009). Modifiable factors associated with changes in postpartum depressive symptoms. *Archives of Women's Health, 12(2)*, 113-120.

Howell, E., Mora, P. A., Horowitz, C. R. & Leventhal, H. (2005) Racial and Ethnic Differences in Factors Associated With Early Postpartum Depressive Symptoms. *Obstetrics & Gynecology, 105*, 1442-1450.

Howell, E., Mora, P. A., & Leventhal, H. (2006). Correlates of Early postpartum depression. *Maternal and Child Health Journal, 10(2)*, 149-157.

Jenkinson, C., Chandola, T., Coulter, A., & Bruster, S. (2001). An assessment of the construct validity of the SF-12 summary scores across ethnic groups. *Journal of Public Health Medicine, 23(2)*, 187-194.

Johnson, J.E. (1973). Effects of accurate expectations about sensations on the sensory and distress components of pain. *Journal of Personality and Social Psychology, 27*, 261-267.

Johnson, J.E., & Leventhal, H. (1974). The effects of accurate expectations and behavioral instructions on reactions during a noxious medical examination. *Journal of Personality and Social Psychology, 29*, 710-718.

Johnson, J. E., Leventhal, H., & Dabbs, Jr., J. M. (1971). Contribution of emotional and instrumental response processes in adaptation to surgery. *Journal of Personality and Social Psychology, 20 (1)*, 55-64.

Johnson, J.E., Morrissey, J. F., & Leventhal, H. (1973). Psychology preparation for an endoscopic examination. *Gastointestinal Endoscopy, 19*, 180-182.

Jones, R.N., & Gallo, J.J. (2002). Education and sex differences in the mini-mental state examination: effects of differential item functioning. *Journal of Gerontology: Psychological Sciences, 57B(6)*, P548-P558.

Katula, J.A., Rejeski, W.J., Wickley, K.L., & Berry, M. J. (2004). Perceived difficulty, importance, and satisfaction with physical function in COPD patients. *Health and Quality of Life Outcomes*, 2, 18. doi:10.1186/1477-7525-2-18.

Kline, C. R., Martin, D. P., Deyo, R. A. (1998). Health consequences of pregnancy and childbirth as perceived by women and clinicians. *Obstetrics & Gynecology, 92*, 842-848.

Klossner, N. J. (2006). *Introductory maternity nursing*. Philadelphia: Lippincott Williams & Wilkins.

Leahy-Warren, P. (2007). Social Support for First-Time mothers: An Irish Study. *The American Journal of Maternal/Child Nursing, 32(6)*, 368-374.

Leventhal, H., Halm E., Horowitz, C., Leventhal, E. & Ozakinci, G. (2004). Living with chronic illness: A contextualized, self-regulation approach. In S. Sutton, A. Baum & M. Johnston (Eds.), The Sage Handbook of Health Psychology, (pp. 197-240), London: Sage.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd Ed.). Hoboken: JN, John Wiley & Sons, Inc.

Littleton, L. Y., & Engebretson, J. (2002). *Maternal, neonatal, and women's health nursing.* Australia: Delmar Thomson Learning.

MacKinnon DP . Introduction to statistical mediation analysis. Mahwah, NJ: Erlbaum; 2008.

Marquez, D. X., Bustamante, E. E., Bock, B. C., Markenson, G., Tovar, A., & Chasan-Taber, L. (2009). Perspectives of Latina and Non-Latina White Women on Barriers and Facilitators to Exercise in Pregnancy. *Women Health, 49(6)*, 505-521. doi:10.1080/03630240903427114.

Mazor, K. M, Hambleton, R. K., & Clauser, B. E. (1998). Multidimensional DIF Analyses: The Effects of Matching on Unidimensional Subtest Scores. *Applied Psychological Measurement, 22(4)*, 357-367.

McDonough, J., Gibbs, B., Scott-Harris, J., Kronebusch, K., Navarro, A., & Taylor, K. A. (2004). State Policy Agenda to Eliminate Racial and Ethnic Health Disparities. Commonwealth Fund.

Medical Outcomes Trust. (2010). Instruments. Retrieved June 9, 2010 from http://www.outcomes-trust.org/instruments.htm.

National Academy of Sciences. (2009). Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement. Washington, DC: The National Academies Press.

RAND Health. (2010). Medical Outcomes Study: Measures of Patient Adherence. Retrieved June 9, 2010, from http://www.rand.org/health/surveys_tools/mos/mos_adherence.html.

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. Quality of Life Research, 16(S1), 19-31, DOI: 10.1007/s11136-007-9183-7.

Steinberg, L., & Thissen, D. (2006). Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. *Psychological Methods, 11*, 402-415.

Stewart, A.L., Ware, J.E., & Brook, R.H. (1978). Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. II, Physical Health in Terms of Functioning. Santa Monica: The RAND Corporation; Pub. No. R-1987/6-HEW.

Stewart, A.L., & Ware, J.E. (1992). *Measuring Functioning and Well-being*. Durham, NC: Duke University Press.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics.* Boston: Allyn and Bacon.

Teresi, J. A. (2006). Overview of Quantitative Measurement Methods: Equivalence, Invariance, and Differential Item Functioning in Health Applications. *Medical Care, 44(11)*, S39-49.

Teresi, J. A., Kleinman, M., Ocepek-Welikson, K. (2000). Modern psychometric methods for detection of differential item functioning: application to cognitive assessment measures. *Statistics in Medicine, 19*, 1651-1683.

Thornton, P.L., Kieffer, E.D., Salabarria-Pena, Y., Odoms-Young, A., Willis, S.K., Kim, H., & Salinas, M.A. (2006). Weight, diet, and physical activity-related beliefs and practices among pregnant and postpartum latino women: the role of social support. *Maternal and Child Health, 10*, 95-104.

Ware, J., Kosinski M., & Keller S.D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical Care, 34*, 220-233.

Webb, D. A., Bloch, J. R., Coyne, J. C., Chung, E. K., Bennett, I. M., & Culhane, J. F. (2008). Postpartum Physical Symptoms in New Mothers: Their Relationship to Functional Limitations and Emotional Well-being. *Birth*, *35*, 3.

Wallace, P. M. & Gotlib, I. H. (1990). Marital Adjustment during the Transition to Parenthood: Stability and Predictors of Change. *Journal of Marriage and Family, 52*(1), 21-29.

Williams, C.D., Taylor, T.R., Makambi, K., Harrell, J., Palmer, J.R., Rosenberg, L., & Adams-Campbell, L.L. (2007). CES-D four-factor structure is confirmed, but not invariant, in a large cohort of African American women. *Psychiatry Research, 150*, 173–180.

Yang, F. M., Tommet, D., & Jones, R. N. (2009). Disparities in self-reported geriatric depressive symptoms due to sociodemographic differences: An extension of the bi-factor item response theory model for use in differential item functioning. *Journal of Psychiatric Research, 43*(12): 1025–1035. doi:10.1016/j.jpsychires.2008.12.007.

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. AAT 3066425).

BIOGRAPHICAL INFORMATION

Tamer Desouky was born in Fairfax, Virginia and spent his childhood in Dhahran, Saudi Arabia's Eastern Province until eight years of age. Afterwards his family moved back to the states and eventually settled in a small town in southwest Virginia. He graduated with a Bachelor of Science degree from Virginia Tech double majoring in Psychology and Sociology. After a brief one year Post-baccalaureate Program in Biology Tamer enrolled in the Ph.D. program in Health Psychology and Neuroscience at the University of Texas at Arlington. His research interests include Research Methodology and Quantitative Methods with substantive areas in Health Psychology and Neuroscience. He plans to complete his PhD, continue working in the field of research methodology, and seek more knowledge while at the same time generating it and imparting it to others.