DIFFERENCES IN DEPRESSIVE SYMPTOM PRESENTATION BETWEEN LATINO AND NON-LATINO WHITE YOUTHS ON THE CENTER FOR EPIDEMIOLOGIC STUDIES DEPRESSION SCALE (CES-D): A MODERATED NONLINEAR FACTOR ANALYSIS (MNLFA) APPROACH

Stephanie Salcedo

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology and Neuroscience (Clinical Psychology).

Chapel Hill
2019

Approved by:

Eric A. Youngstrom

Anna Bardone-Cone

Patrick J. Curran

Andrea M. Hussong

Deborah J. Jones

ABSTRACT

Stephanie Salcedo: Differences in Depressive Symptom Presentation between Latino and non-Latino White Youths on the Center for Epidemiologic Studies Depression Scale (CES-D): A Moderated Nonlinear Factor Analysis (MNLFA) Approach
(Under the direction of Eric A. Youngstrom)

Most mental health measures have been validated with English speaking and majority White samples (Guillemin et al., 1993; Vega & Rumbaut, 1991). Despite guidelines recommending formal cultural adaptation to ensure that scale content still measures what it intended and provide accurate group comparisons, such adaption is rare (Borsboom, 2006; Chen, 2008). Latinos are the largest minority in the US, making it imperative to understand how cultural factors can influence mood symptom endorsement and conceptualization. This study examined the Center for Epidemiologic Studies Depression Scale (CES-D) for differential item functioning (DIF). Secondary analyses of Latino (*n*=3,208) and non-Latino White youths (*n*=9,919) from the National Longitudinal Study of Adolescent Health (Add Health). Moderated nonlinear factor analysis (MNLFA) examined ethnicity DIF, while accounting for other covariates (gender, age, parental education, acculturation). An exploratory aim compared Latino subgroups separately (Mexicans, Puerto Ricans, and Cubans) to assess subgroup differences in the item intercepts and factor loadings. In the initial model (only incorporating ethnicity), at similar levels of depression, Latinos were more likely to score higher on being bothered by things, feeling unable to shake the blues, talking less, and feeling like a failure. However, Latinos were less likely to endorse having sleep problems, feeling like others were unfriendly, having crying spells, or difficulty getting started on things. There were also significant differences in the

factor loadings for the *blues* and *crying* items. In the second model incorporating covariates (e.g., gender, age, parental education), nine items showed Latino-White DIF (*ps*<.05), but two items (*being bothered by things, crying spells*) no longer showed ethnicity DIF. However, graphical analyses showed that DIF effects on the overall measure appear small and unlikely to affect total scores between groups. Given that the magnitude of DIF observed on the CES-D between Latinos and non-Latinos in this sample was small, the CES-D measured depressive symptoms similarly between Latino and White youths. Using MNLFA to examine DIF provided a more comprehensive picture of how demographic characteristics influenced symptom reporting. MNLFA modeled covariates simultaneously when assessing DIF, avoiding potentially confounding that might otherwise cloud our understanding of cultural differences.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CES-D   Center for Epidemiologic Studies Depression Scale

CDI   Child Depression Inventory

CFA   Confirmatory Factor Analysis

DIF   Differential Item Functioning

DSM-5   Diagnostic and Statistical Manual for Mental Disorders, Fifth Edition

EFA   Exploratory Factor Analysis

ICC   Item Characteristic Curve

IDA   Integrative Data Analysis

IRT   Item Response Theory

MI   Measurement Invariance

MIMIC   Multiple-Indicator Multiple Cause

MNLFA   Moderated Non-Linear Factor Analysis

POMP   Percentage of Maximum Possible

SES   Socioeconomic Status

INTRODUCTION

Depressive disorders affect a significant number of individuals across the lifespan; some studies estimate that major depressive disorder affects 1% of children, then the rates rise to 4-5% in adolescence (Costello, Egger, & Angold, 2005; Kessler, Avenevoli, & Ries Merikangas, 2001). Having depression at a younger age is predictive of future impairment from depression in adulthood (Birmaher, Arbelaez, & Brent, 2002). Adolescents with depression demonstrate poorer academic performance (with increased school dropout rates), interpersonal relationship problems, and more negative self-esteem (Garber & Horowitz, 2002; Hammen & Rudolph, 2003; Lewinsohn & Essau, 2002; Waslick, Kandel, & Kakouros, 2002). The need for early intervention is crucial because depression is associated with an increased risk of suicide (Nock & Kessler, 2006) and the majority who experience their first episode have another one later in their lives (Kessing, Hansen, & Andersen, 2004). Despite the negative consequences of untreated depression, it often takes many years before an individual is diagnosed and treated for the illness (Kessler et al., 2003; Kessler, Merikangas, & Wang, 2007). Therefore, more work needs to be done to identify obstacles impeding accurate assessment and treatment, given the association between earlier onset of depression and impairment from depression in adulthood (Birmaher et al., 2002).

Evidence-Based Measures to Improve Assessment Accuracy in Depression

Traditionally, clinicians have used unstructured interviews and their clinical judgment to determine an individuals' diagnosis. However, one problem with unstructured interviews is that clinicians may not ask about all the necessary diagnostic criteria or overlook other potential

1

diagnoses, potentially leading to inaccurate clinical impressions (Kessler et al., 2001; Spitzer, 1983). Therefore, semi-structured clinical interviews are considered the "gold standard" to assess for disorders such as depression because they provide a guide for clinicians to follow and ask the questions needed to assess diagnostic criteria; imposing a structure and incorporating with clinical judgment reduced the likelihood of engaging in various biases, such as selectively incorporating information that confirms the clinician's initial impressions and overlooking contrary evidence (Angold, 2002). Another benefit of using semi-structured interviews is that clinicians have the flexibility to ask more follow-up questions that go beyond the standardized interview questions, which could contribute to a more comprehensive picture of the client's clinical presentation.

Rating scales and checklists are another way to assess mood symptoms and can be attractive to clinical settings because they do not require extensive time or training to administer and can have high face validity especially when questions match specific diagnostic criteria (Jensen & Haynes, 1986). Individuals can rate the presence and/or severity of particular symptoms, which can then be converted to overall summary scores that can inform the clinician of the client's presenting concerns and degree of severity or impairment. Furthermore, individuals' scores can be normed to compare them with others in their age range to determine the degree to which their symptoms deviate from others similar to them. When assessing children and adolescents for depression, however, incorporating rating scales completed by their parents can be valuable because youths may not have the insight into their symptoms, may not feel comfortable disclosing their mood symptoms, or may have trouble recalling specific information about the onset, frequency, or duration of their symptoms (Garber & Kaminski, 2000). The added value of parent report can depend on the age of the youth. Young children are

less reliable reporters of their psychological symptoms than adolescents (Edelbrock, Costello, Dulcan, Kalas, & Conover, 1985), and parents are likely more involved in the everyday lives of their children than of their adolescents, making them better able to provide specific details about the symptoms they have observed (Richters & Pellegrini, 1989).

Latinos and Depression

Latinos are the largest minority group in the U.S., with 56.5 million people (17.6% of national population; U.S. Census Bureau, 2015), so more attention has been directed to assessing their mental health outcomes. Research on racial/ethnic differences in depression prevalence rates between Latinos and non-Latino Whites has yielded mixed findings. Some researchers have found that Latinos had lower rates of lifetime mood disorders (Kessler et al., 2005a), whereas other studies found no differences (Kessler, Chiu, Demler, & Walters, 2005b; Kubik, Lytle, Birnbaum, Murray, & Perry, 2003). Some studies in youths have found higher depressive symptom levels in Latinos (e.g., Allen & Astuto, 2009; Paxton, Valois, Watkins, Huebner, & Drane, 2007). One possibility for these differences in prevalence rates is whether depression assessments function differently cross-culturally.

In the psychological literature, the terms *Hispanic* and *Latino* are often used interchangeably, but some argue they differ in meaning. The term *Hispanic* was created in the 1970s by the U.S. Census Bureau to categorize individuals of Spanish heritage (Ennis, Rios-Vargas, & Albert, 2011). However, because the term highlights European origins as their common characteristic and overlooks indigenous cultures (Quiñones-Rosado, 2002), others have preferred the use of the term *Latino*, which refers to people of Latin American heritage, taking into account indigenous and African influences in addition to their common history of Spanish colonization in their host countries (Ramos & Magana, 2008). The terms *Hispanic* and *Latino*

3

will be used interchangeably to be consistent with the terms used in previous studies. For individuals who identify as being of European descent, the terms *White, non-Latino/Hispanic White,* or *Anglo* will be used, consistent with the cited authors' choice of terms.

The Importance of Cross-Cultural Adaptation

Most measures used in research have been developed and validated in English-speaking countries and with predominantly White samples (Guillemin, Bombardier, & Beaton, 1993; Vega & Rumbaut, 1991). Therefore, to use a measure in a different language or with a different cultural group, a cultural adaptation process is recommended to make sure that the scale's content it was intended to measure stays preserved, and the measure scores can be used to make accurate group comparisons (Borsboom, 2006; Chen, 2008; Ferraz, 1997; Vaughn-Coaxum, Mair, & Weisz, 2016). For example, if one group scores higher on a depression measure than another group for reasons other than their depressive severity, then measure is biased, and examining group mean differences can be misleading. In addition, if clinical cut-off scores were established based on the scores of one group, bias can lead to inaccurate estimates in prevalence rates of a different group (Crockett, Randall, Shen, Russell, & Driscoll, 2005).

The goal is to adapt the measure to ensure the new scale maintains semantic, idiomatic, experiential, and conceptual equivalence (Beaton, Bombardier, Guillemin, & Ferraz, 2000). *Semantic equivalence* is when the meaning of the words or phrases is preserved after translation and/or cultural adaptation. *Conceptual equivalence* is when a construct has a same meaning across groups, because an item may translate but still not be conceptually equivalent (Beaton et al., 2000). One example is the meaning of the word "family." In the U.S., individuals may think of their immediate relatives (e.g., if a parent, may think of spouse and children). In other countries, such as those in Latin America, individuals may think of their extended relatives as

well (e.g., cousins, aunts, siblings). *Idiomatic equivalence* is when colloquialisms or idioms are adequately translated to conserve their meanings. For instance, "feeling blue" means feeling sad or down in English, but this phrase would not hold its meaning if directly translated to Spanish. *Experiential equivalence* is when a situation or context described in the measure is adjusted (if needed) to fit the new cultural context. For example, if a question asks about ability to being able to drive a car, one might need to adjust the item if asking the question in a culture where most people do not have access to a car (Guillemin et al., 1993).

Guidelines for the adaptation of self-report measures have been recommended, with different outlined steps depending on the culture, language, and country of use intended (Beaton et al., 2000; Guillemin et al., 1993). Although it may seem intuitive to conduct a formal adaptation process when translating a measure in a new language, the guidelines still recommend that a cultural adaptation take place even when using the measure in the same country and language but with a different culture, because if measures are not equivalent, it is not possible to make valid group comparisons (Hambleton & Kanjee, 1995; Vandenberg & Lance, 2007). Analyses to assess the equivalence of measures can include item level (e.g., item-to-scale correlations, internal consistency) and score-level characteristics (e.g., reliability, construct validity; Beaton et al., 2000). Differential item functioning is one way to assess *measurement invariance* (Millsap, 2012).

Measurement Invariance and Differential Item Functioning: An Overview

Measurement invariance (MI) can be defined as the assumption that the scale in question is measuring the same construct across groups (Borsboom, 2006; Chen, 2008; Vaughn-Coaxum et al., 2016). If there is bias in the measure, it can confound the scores, making the group mean differences misleading. In addition, for mental health measures, using a clinical cut-off score that

was established for one group may lead to under- or over-estimates in prevalence estimates in a different group (Crockett et al., 2005). In the Item Response Theory (IRT) literature, the lack of measurement invariance is called differential item functioning (DIF). DIF occurs when specific items in a measure show statistically significant differences in responses across groups after controlling for the specific level of the construct being measured, such as depression, for instance (Holland & Wainer, 1993). When DIF is present, individuals' trait severity level alone does not account for their individual responses (De Ayala, 2008). For example, if one is examining responses on a depression scale between two different ethnic groups, DIF is present when at similar depressive severity levels, one group has a higher probability of endorsing an item than the other group.

To ensure MI, *configural invariance* is necessary (though not sufficient); this means that the items comprising the measure show a similar factor structure between groups (Horn & McArdle, 1992; Meredith, 1993). In other words, the scale indicators correspond to the same underlying concepts for each of the groups; the same item needs to be associated with the same latent factor in each group, but the factor loadings can differ (Horn, McArdle, & Mason, 1983). Configural invariance suggests that in a depression measure, the groups being compared share the same concept of depression, and lack of invariance would mean that a different set of items might be needed to assess for a particular factor in a different group (Pina, Gonzales, Holly, Zerr, & Wynne, 2013).

Examining other aspects related to item equivalence can be more useful for understanding clinically and culturally relevant differences in groups. One of these is *metric invariance,* which occurs when the item factor loadings in a measure are similar across groups (Labouvie & Ruetsch, 1995); differences in factor loadings could mean there are differences in

the clinical presentation or in the interpretation of an item, or that some items are better indicators of the symptom cluster in one group than another. Factor loadings can also be conceptualized as slopes of lines regressing item responses on the latent factor, so when the factor loadings are equal, their unit of measurement is identical (Bollen, 1989). This is analogous to non-uniform DIF, which occurs when the probability of one group endorsing a particular item more than the other group differs depending on the trait level, similar to an interaction effect. *Scalar invariance* is also important to consider, which is when the intercepts of the items across groups are similar. A violation of scalar invariance is also known as uniform DIF, which occurs when the probability of endorsing a particular item is consistently higher in one group across trait severity levels, so there is group difference in the probability of endorsing an item at all levels of the latent trait (Crane, van Belle, & Larson, 2004; Mellenbergh, 1982). Looking beyond configural invariance can be more informative because one can gather information about the nuances of symptom presentation and how they differ by group characteristics, which could then inform diagnostic impressions and treatment planning (Pina et al., 2013).

The degree of DIF is known as the *magnitude*, which can be measured using effect size statistics, such as odds ratios, beta coefficients, and change in *R*-square (Teresi, Ramirez, Lai, & Silver, 2008). Because DIF is influenced by the sample size, assessing the magnitude can help to elucidate whether the DIF is clinically meaningful. Even if items show DIF, the effect may be small enough to not affect the overall scores. To look beyond the item level, the impact of DIF on the measure can be examined by assessing how means of a measure change with and without the inclusion of the items with DIF; when cut-off scores are established for a measure, one could also determine if cut-scores should be adjusted (Stark, Chernyshenko, & Drasgow, 2004). Experts caution against merely deleting items that show significant DIF because they can be

culturally informative for understanding differences in symptom presentation (Knight, Tein, Prost, & Gonzales, 2002; Pina, Little, Knight, & Silverman, 2009). For example, Knight et al. (2002) illustrate that a suicide item on a depression scale may load on factors differently by ethnic group because of cultural differences in the acceptability of endorsing suicidal thoughts, leading to a more limited range of answers. However, deleting this question just to achieve measurement invariance would cause a loss of important clinical and cultural information that could inform future efforts to better understand differences in symptom presentation and improve their detection.

Factors that Can Lead to Measurement Bias

Identifying possible factors that can lead to differential item response across race/ethnicity can help to better adapt measures to serve diverse communities. There are several factors related to Latinos that may influence how they conceptualize their mental health symptoms, which in turn, can influence how they answer questions on a depression scale. However, more work needs to be done on factors that can influence within group variation. Simply looking at differences between Latinos and Whites is likely masking other effects related to other demographic variables. Latinos come from over 20 countries, with each country having its own unique cultural and historical traditions (Cauce & Domenech-Rodriguez, 2002). Mexicans (66% of Hispanics), Puerto Ricans (9%), Cubans (4%), and Salvadorans (4%) are the largest Latino subgroups, and there are vast differences in characteristics including their socioeconomic status, level of education, acculturation, and immigrant journeys (Alegria & Woo, 2009; Motel & Patten, 2012). In addition, Latinos tend to keep ties with friends and loved ones in their home countries, which helps to maintain their language and unique cultural traditions despite living in the U.S. (Levitt, DeWind, & Vertovec, 2006; Viruell-Fuentes, 2006).

Therefore, measurement invariance research should move beyond solely focusing on Latino-White differences and instead incorporate other explanatory variables that can help uncover sources of response bias (Pina et al., 2013). Indeed, research has shown that including background characteristics in psychological models improves the score quality relative to models that did not incorporate covariates (Curran, Cole, Bauer, Hussong, & Gottfredson, 2016; Curran, Cole, Bauer, Rothenberg, & Hussong, 2018). Below we describe some of these characteristics, categorized by cultural and demographic factors, and how they might influence between- and within-group variation in depressive symptom expression.

Cultural Factors.

Somatization of psychological distress. Research suggests that Latinos are among some of the minority groups who are more likely to *somaticize* mental health symptoms, which means to express their psychological distress in the form of physical ailments (e.g., insomnia, fatigue, appetite loss, restlessness); indeed, the most recent iteration of the Diagnostic and Statistical Manual for Mental Disorders (DSM-5) states that "in many cultures, somatic symptoms are very likely to constitute the presenting complaint" (American Psychiatric Association, 2013). Depression assessment guidelines recommend that when working with Latinos, assessing somatic symptoms is important because they can reveal underlying depressive symptoms (Lewis-Fernández, Das, Alfonso, Weissman, & Olfson, 2005). Previous studies in adults and adolescents have found that Latinos tend to endorse more physical symptoms for their mental distress (Canino, Rubio-Stipec, Canino, & Escobar, 1992; Choi & Park, 2006; Escobar, Burnam, Karno, Forsythe, & Golding, 1987; Kolody, Vega, Meinhardt, & Bensussen, 1986; Roberts & Sobhan, 1992).

However, other researchers argue that rates of somatization of mental health symptoms are similar cross-culturally. Kirmayer (2001) argued that somatization rates depend greatly on the medical setting; for instance, in a primary care setting, patients with depressive symptoms may feel like describing their somatic complaints is more relevant given that they are seeing medical doctors, not mental health specialists. Therefore, comparing somatization rates without taking into account the setting may lead to false conclusions about cultural differences (Kirmayer, 2001). In line with this argument, previous studies found no cultural differences in rates of somatization of depressive and anxiety symptoms in primary care settings (Kirmayer, Robbins, Dworkind, & Yaffe, 1993; Kroenke et al., 1997; Simon, VonKorff, Piccinelli, Fullerton, & Ormel, 1999). However, Latinos may be less likely to express negative emotions due to increased stigma about mental illness (Varela, Weems, Berman, Hensley, & De Bernal, 2007). Therefore, if Latinos do choose to describe their symptoms, they may feel more comfortable describing their physical or somatic complaints because they are less stigmatizing (Epstein, Quill, & McWhinney, 1999). One way to examine the rates of somatization between Latinos and Whites is to compare item responses to somatization questions in depression measures; if Latinos are more likely to endorse somatic symptoms when experiencing depression, then there would be DIF on these items because when accounting for overall depressive severity level, Latinos would be more likely to endorse somatic symptoms as compared to Whites. In addition, we might also observe differences in the factor loading of the somatic symptoms if somatic symptoms are more core to depression as a factor structure in Latinos than in Whites.

Acculturation. Acculturation is the physical and psychological process of adapting to living in a new place with different customs or norms (Wells, Golding, Hough, Burnam, &

10

Karno, 1989). At the psychological level, these processes can include the incorporation of new cultural beliefs, values, behaviors, as well as learning a new language and gaining a sense of belonging (Rogler, Cortes, & Malgady, 1991). Acculturation is crucial to the study of mental health in Latinos because of its influence on mental health related factors. More specifically, *acculturative stress* can result from the process of acculturation, which can include exposure to stressful life events, language problems, perceived discrimination, or feeling one's cultural values are incompatible with the host culture's (Vega, Zimmerman, Gil, Warheit, & Apospori, 1997). The link between acculturation and mental health outcomes, however, is not clear, with some finding higher rates among U.S.-born Latinos, while others showing greater psychopathology among immigrant groups (Alegría et al., 2008; Bas-Sarmiento, Saucedo-Moreno, Fernández-Gutiérrez, & Poza-Méndez, 2017). One possibility for these discrepant findings is that how acculturation is measured varies vastly by study (Alegria & Woo, 2009). Experts recommend examining multiple domains such as language use, food preferences, ethnic identity, and alignment with cultural values (Cuellar, Arnold, & Maldonado, 1995), as well as also looking at retention of one's own culture (Berry, 2003; Cabassa, 2003).

Most of the acculturation research has been conducted in adults, and the processes in youths may differ because of their varied social contexts (school, peers, family dynamics; Oetting, Donnermeyer, Trimble, & Beauvais, 1998). Youths are still developing cognitively, socially, and emotionally, so they may be more vulnerable to stressors related to acculturation (Gonzales, Knight, Morgan-Lopez, Saenz, & Sirolli, 2002). Previous research has found no relationship between acculturation levels and depressive symptoms in Mexican youths, but there was a positive correlation between acculturative stress and depression (Hovey & King, 1996; Katragadda & Tidwell, 1998). Other studies have also come to the opposite conclusions, finding

that higher acculturation (as measured by English competence) was associated with lower depressive symptoms (Rumbaut, 1996). The most conclusive findings related to acculturation in Latino youths has been the link between greater acculturation and problem behaviors such as substance use and externalizing problems, and authors theorize that it could be related to Latino youths' increased vulnerability to discrimination pressures, minority status, increased exposure to deviant peers when at school, and decreased adherence to family cultural values (Gonzales et al., 2002; Yi, Chen, Hussong, & Daughters, 2016). Therefore, more research is needed to understand what facets of acculturation play a larger role in the presence of mental health symptoms, particularly depression.

There is a line of research that has examined acculturation's role in Latino individuals' responses to self-report questionnaires. Marin et al. (1992) found that Hispanics were more likely to endorse extreme responses on a Likert scale and agree more with a given item than non-Hispanic Whites, consistent with previous findings (Hui & Triandis, 1989). However, with *increased* acculturation, Hispanics were *less* likely to show extreme responses and agree with the items *less* frequently. In addition, this effect still held after controlling for education (Marin et al., 1992). A value Latinos may have been socialized to is *simpatía*, which encourages individuals to have smooth interpersonal relationships and be more collectivistic (Booth-Kewley, Rosenfeld, & Edwards, 1992). How alignment with simpatía may influence response is that Latinos may choose to rate themselves higher on self-report measures because that is indicative of showing genuine, unmoderated feelings, allowing for increased openness and responsiveness to the group's needs (Hui & Triandis, 1989). Therefore, Latinos who are less acculturated to U.S. culture likely feel more connected to their own culture and thus are more likely to value simpatía. Hopwood et al. (2009) expanded on this research and found that Latinos scored

significantly higher than Anglo individuals on socially desirable indicators, suggesting that there may be cultural differences in how individuals respond to present themselves in a more positive light.

The following questions still remain in this area of research: Are Latinos more likely to agree with or pick more extreme ratings for depressive items as compared to Whites? Is this effect likely to be similar in youths? Are Latinos likely to over-endorse items relating to positive affect to respond in a more socially desirable way? Only two studies have examined the role of acculturation on DIF in a depression measure among Hispanic women. Nguyen, Clark, and Ruiz (2007) found that the "non-acculturated" Hispanic pregnant women group (defined as preference for Spanish CES-D) was less likely to endorse somatic symptom items but more likely to endorse positive affect items than the Hispanic group that preferred the English CES-D. However, it is unclear whether the DIF is related to acculturation or to the translation of the measure. McCabe and colleagues (2011) found that when comparing Spanish CES-D responses between Hispanic women with *high* versus *low* acculturation levels (as defined by an acculturation measure), DIF was only present in one item; women with low acculturation more strongly endorsed that "people were unfriendly." One possibility for this finding is that individuals who are less acculturated may face more discrimination in their everyday lives. Although these studies examine different aspects of acculturation, they highlight the need for future studies to examine acculturation using multiple questions or acculturation measures, because uncovering the sources of DIF requires being able to tease apart what differences are related to the translation of a measure versus how assimilated one feels to the host culture. Therefore, for future studies, examining self-report acculturation as a covariate in DIF studies could elucidate within the Latino group the source of the DIF.

Demographic Factors.

Gender. Previous findings show a higher prevalence of depressive symptoms in girls relative to boys starting in adolescence (Hankin et al., 1998). Several theories explaining the increase of depression among girls have included interactions between stress and biological changes associated with puberty (Cyranowski, Frank, Young, & Shear, 2000), girls' increased exposure to interpersonal challenges (Shih, Eberhart, Hammen, & Brennan, 2006), greater risk of exposure to traumatic sexual abuse (Hilt & Nolen-Hoeksema, 2009), as well as gender differences in responses to stressful negative events, with girls tending to adapt a more internalizing and ruminative coping style (Nolen-Hoeksema, 2000).

Another possibility is that boys and girls differ in their endorsement of their symptoms and/or how they respond to questions on measures. Previous research on gender DIF have found differences in how boys and girls respond to items. In a study examining the Child Depression Inventory (CDI) in a sample of 4,000 school-age children and adolescents, Van Beek and colleagues (2012) found that girls were more likely to endorse items examining sadness and crying, at similar depression levels, as compared to boys. In addition, girls endorsed the following symptoms at higher levels and were more characteristic of depression: worrying about the future, self-blaming, and feeling like things bother them all the time (Van Beek, Hessen, Hutteman, Verhulp, & Van Leuven, 2012). This suggests that there may be other symptoms boys endorse when feeling depressed. One possibility is that boys would be more likely to endorsing physical or externalizing symptoms at similar depression levels, but this hypothesis was not supported (Van Beek et al., 2012).

In the MI literature, there is scant research on the interaction between gender and ethnicity in depressive symptom endorsement. Uniquely, gender role values that Latinos may

have been exposed to growing up are *Machismo* and *Marianismo*. *Machismo* is the idea that the man should be the provider, protector, and head of his family (Arciniega, Anderson, Tovar-Blank, & Tracey, 2008). The counterpart to Machismo is *Marianismo*, the belief that women should be self-sacrificing, nurturing, and a source of spiritual strength for their families, characteristics similar to the Virgin Mary, a Catholic icon (Niemann, 2004). Ascribing to these beliefs are associated with higher levels of depression, anxiety, and anger (Fragoso & Kashubeck, 2000; Kopper & Epperson, 1996; Pina-Watson, Castillo, Ojeda, & Rodriguez, 2013; Syzdek & Addis, 2010). Latina girls are socialized to be more emotionally expressive than Latino boys (Vazquez-Nuttall, Romero-Garcia, & de Leon, 1987), so Latino boys may feel less comfortable expressing emotional problems and instead express them as externalizing behaviors (Umaña-Taylor & Updegraff, 2007). However, endorsement of these gender roles has decreased across generations, becoming less rigid and more egalitarian with increased urbanization, migration, and industrialization (Hurtado, 1992; Ojeda, Rosales, & Good, 2008; Phinney & Flores, 2002). These are values that can inform behavior, not necessarily behavior patterns themselves (Cauce & Domenech-Rodriguez, 2002). Therefore, if Latina girls were socialized about Marianismo believe in these norms, then they may under-endorse items that could conflict with their beliefs that they should be strong and self-sacrificing. Similarly, Latino men may be more hesitant to endorse emotional symptoms of depression (e.g., feeling sad, worthless) because they may think that expressing emotions conflict with masculinity.

 Age. Given that children's ability to have insight about their experiences and understand their mental states develops as they get older, children's endorsement of depression symptoms may differ depending on their age. Depression rates increase around age 14 in adolescence, particularly in girls (Hankin et al., 1998). Van Beek et al. (2012) found numerous instances of

measurement invariance between adolescents and younger children; for instance, questions referring to self-esteem described the overall factor Self-Deprecation better for participants in adolescence as compared to those in elementary school. They also reported that at lower overall levels on the School Problems factor relative to elementary school children, early adolescent participants reported more difficulties with schoolwork (Van Beek et al., 2012). To the best of our knowledge, no previous study has examined whether similar findings would be observed in Latino youths.

Socioeconomic status. Vast disparities exist between Latinos and Whites on income and education. Latinos are less likely to have a high school or college degree, less likely to have health insurance, and be of a lower income bracket as compared to Whites (Marotta & Garcia, 2003). Individuals of a lower socioeconomic status may face greater stressors such as finding work, having enough money to pay bills, risks of living in unsafe neighborhoods or unstable housing, and lack of resources to seek medical or mental health treatment when needed. Being of a lower SES has been associated with increased depression and anxiety, among other mental health problems (Conger, Ge, Elder, Lorenz, & Simons, 1994; Goodman, Slap, & Huang, 2003). Research on the effects of SES on DIF in depression measures is scant; however, given that Latinos and Whites may differ in SES, it is important to consider whether any ethnic differences observed in item response remain after accounting for differences in education level and family income.

Center for Epidemiologic Studies Depression Scale (CES-D)

The CES-D is a 20-item self-report inventory that assesses frequency of depression symptoms over the past week (Radloff, 1977). The CES-D was originally developed to measure somatic and affective symptoms in in adult community samples but has been validated for use

with youths and adolescents (McArdle, Johnson, Hishinuma, Miyamoto, & Andrade, 2001; Weissman, Orvaschel, & Padian, 1980). Radloff (1977) found support for the following four-factor structure of the CES-D: positive affect, negative affect, psychosomatic complaints, and interpersonal problems. There has been extensive debate in the literature as to whether this four-factor solution applies to Hispanics and other non-White samples (e.g., Liang, Tran, Krause, & Markides, 1989; Miller, Markides, & Black, 1997; Stroup-Benham, Lawrence, & Treviño, 1992). Kim et al. (2011) conducted a meta-analysis of 28 studies examining racial/ethnic differences in the CES-D factor structure, summarizing the findings by those using exploratory factor analysis (EFA), a data-driven approach to determine the best-fitting underlying structure, and confirmatory factor analysis (CFA), a model-driven approach using theory or previous results. They found that for the CFA studies, Radloff's (1977) four-factor structure was replicated in most racial/ethnic groups (White, African American, Hispanic, American Indian) except for Asians. However, in the EFA studies, the CES-D factors differed greatly by racial/ethnic groups; in particular, Hispanics tended to show four factors, but the factor loadings of depressed affect and somatic symptom items sometimes switched between the two factors (Kim et al., 2011). Although these findings are mixed, they highlight the possibility that cultural differences in the conceptualization, meaning, and symptom expression of depression exist. One thing to note, however, is most of these studies were with adults. Two previous studies in school-aged youths also found that the four-factor model did not fit well across Latinos (Crockett et al., 2005; Perreira, Deeb-Sossa, Harris, & Bollen, 2005).

Given that most studies using the CES-D sum all the items for a total score, rather than looking at separate factor scores, Edwards and colleagues (2010) examined whether a one-factor model would provide a good fit. They found that a one-factor model—removing the four

reverse-scored positive affect items—provided a good fit; a two factor model in which these four positive affect items were in a separate factor also showed comparable fit (Edwards, Cheavens, Heiy, & Cukrowicz, 2010). More research is needed to determine whether this factor structure would hold for other racial/ethnic groups or for youths.

Previous DIF findings with the CES-D in Latinos. Only four studies have examined uniform DIF with Hispanics; however, none of the samples have included youths. These studies found that in aggregate, Hispanics were *more* likely to endorse particular somatic symptoms and depressed affect items, but the significant items, for the most part, differed by study (Boutin-Foster, 2008; Iwata, Turner, & Lloyd, 2002; Kim, Chiriboga, & Jang, 2009; MacIntosh & Strickland, 2010). The findings for the positive affect and interpersonal problems were even more mixed, with two studies finding that Hispanics were *less* likely to endorse certain items in these factors (Boutin-Foster, 2008; Iwata et al., 2002; MacIntosh & Strickland, 2010), while others finding the opposite pattern (Kim et al., 2009). Regarding non-uniform DIF, only one study examined youths. Crockett et al. (2005) found that the factor loadings for three of the 20 items were significantly higher for Mexican American youths than for Anglo youths ("felt life was a failure," "felt fearful," and "enjoyed life"), meaning that these symptoms might be more central indicators of the particular factor within depression in Mexican Americans than in Anglo youths.

There are several possible factors that could account for these conflicting findings. Only two studies controlled for demographic variables (e.g., age, gender, and/or education; Boutin-Foster, 2008; MacIntosh & Strickland, 2010). In addition, three studies examined elderly adults (Boutin-Foster, 2008; Kim et al., 2009; MacIntosh & Strickland, 2010), whereas two focused on young adults (Crockett et al., 2005; Iwata et al., 2002). Furthermore, although most used partial

18

correlations to assess for DIF (Boutin-Foster, 2008; Iwata & Buka, 2002; MacIntosh & Strickland, 2010), they differed in the way they calculated sum scores used to correlate items in question. One assesses the group by item partial correlation by conditioning it on the sum score of the remaining items; when an individual's item response is associated with their group membership after accounting for severity level, it indicates that uniform DIF is present (Stricker, 1982). MacIntosh & Strickland (2010) and Iwata et al. (2002) both excluded the positively worded items in their CES-D sum scores due to previous research that Hispanics may have more attenuated positive affect, but Boutin-Foster (2008) included these items, making comparison of findings challenging because they are not using the same comparison group for the correlations. This partial correlation method also has considerable limitations, including not being able to test for non-uniform DIF, higher Type I error rates, and high sampling error with smaller sample sizes (e.g., fewer than 300 in each group; Stricker, 1984). In contrast, Kim et al. (2009) used both IRT and CFA to examine DIF and only identified items that showed DIF under both models, which allows for more conservative estimates, reducing the likelihood of Type I error (Stark, Chernyshenko, & Drasgow, 2006). Overall, previous DIF studies on CES-D have led to inconsistent results on differences in symptom presentation in Latinos, so more research needs to be done to determine the relative roles other cultural factors have on influencing item response.

Moderated Nonlinear Factor Analysis (MNLFA) for DIF/MI

One technique that has not been previously used to assess race/ethnicity DIF in depression symptom ratings is moderated non-linear factor analysis (MNLFA; Bauer & Hussong, 2009). This approach grew out of the integrative data analysis (IDA) approach, which provides recommendations for how to harmoniously combine and analyze data from multiple independent studies (Curran & Hussong, 2009). Bauer (2017) described the benefits of using

19

MNLFA for DIF/MI analyses over other existing analytic approaches, including the Multiple Groups (MG) and multiple-indicator multiple cause (MIMIC) models. Unlike MG models, MNLFA can test for DIF/MI as a function of multiple individual characteristics, including continuous variables, such as age or socioeconomic status, meaning one would not have to dichotomize these variables. In addition, an advantage MNLFA has over MIMIC models is that not only the factor loadings and intercepts but also the variance-covariance parameters can depend on the predictors (Bauer, 2017). Therefore, MNLFA takes the strengths of each of these models to 1) allow all parameters (i.e., factor loadings, variances, covariances) to differ as a function of individual characteristics and 2) incorporate multiple individual characteristics not having to be of the same scale type (e.g., continuous, binary) to assess for DIF simultaneously (Bauer & Hussong, 2009). Using this approach could expand our understanding of DIF because one could test simultaneously whether a depression measure is invariant across ethnicity, age, gender, as well incorporate other important covariates, like acculturation and SES.

Furthermore, an added benefit of estimated MNLFA models is the ability to compute individual-specific factor scores, also known as *modal a posteriori* [MAP] estimates (Bock & Aiken, 1981), which take into account each participant's response pattern, estimates for symptom severity and discrimination, as well as any significant DIF present in the final MNLFA model (Bauer & Hussong, 2009; Cole, Gottfredson, Giordano, & Janssen, 2018; Curran et al., 2014). Advantages over traditional scoring methods also include the ability to have items weighted differently towards the overall score, account for the multidimensional nature of a measure, as well as take individual differences of the sample into account (Gottfredson et al., 2018; Millsap & Everson, 1993).

The Present Study & Hypotheses

The purpose of the present study is to address previous limitations in our understanding of the role that culture plays in depression symptom presentation between Latino and White youths. Beyond solely looking at between-group comparisons in symptom response on the CES-D, one of the most widely used depression measures, we also examined factors that could explain within-group differences in symptom expression, to gain a better understanding of a multitude of factors can lead to DIF. This study had several aims. Aim 1 examined DIF on the CES-D between Latinos and non-Latino White youths. We hypothesized that Latinos would be more likely to endorse the somatic symptom items at similar levels of depression than non-Latino Whites. We also hypothesized that they would show higher factor loadings with the somatic symptoms than non-Latino Whites, suggesting that these symptoms are more central to depression symptom presentation than in Whites. Aim 2 determined whether ethnicity DIF in the CES-D would still be present after accounting for other important covariates (age, gender, SES). We hypothesized that a MNLFA model that incorporates these demographic covariates would result in more accurate factor scores than a model that only incorporated ethnicity because items that show ethnicity DIF may be better explained by other covariates not previously incorporated. Aim 3 examined the interactions between gender and ethnicity to determine whether DIF changes depending on the combinations of these factors. We predicted that there would be an ethnicity by gender interaction, such that Latino boys would be significantly more likely to endorse somatic symptoms than Latino girls, and this gender difference would be significant but less pronounced between White youths. An exploratory aim compared Latino subgroups separately (Mexicans, Puerto Ricans, and Cubans) to examine whether there were any subgroup differences in the intercepts and factor loadings. We also assessed whether there was

acculturation DIF and predicted that Latinos who were less acculturated to U.S. culture will endorse higher somatic symptoms and be more likely to endorse the positively worded CES-D items.

METHODS

Participants and Procedures

These secondary analyses utilized data from the National Longitudinal Study of Adolescent Health, "Add Health," which is a nationally representative sample of adolescents from Grades 7 through 12 (Harris & Udry, 2008). The purpose of this study was to examine factors influencing adolescent health, and the first two Waves focused on factors such as personal traits, family dynamics, interpersonal relationships, schools, neighborhoods, and their communities. Participants were selected with unequal probability from 80 high schools and 52 middle schools, with the intention to be representative of U.S. schools in regard to ethnicity, region, school type and size, and urbanicity. Some minority groups were sampled in proportion to their U.S. population size, but smaller groups, including Puerto Rican and Cuban youths, were over sampled, which makes it possible to conduct more within ethnic group analyses (Harris & Udry, 2008). Add Health is the largest and most comprehensive longitudinal survey conducted on adolescents, with adolescents being interviewed in 1995, 1996, 2001-2002, 2008, and 2016. The study began as in-school questionnaires, and students who completed these or were from one of the participating schools were eligible to complete a home interview ($N = 20,745$ completed in-home surveys). Our analyses used Wave I adolescents who were in Grades 7-12 during the 1994-1995 school year and who did not have missing CES-D data on all variables.

Measures

Demographic Variables. Regarding race/ethnicity, participants were included in the non-Hispanic White group if they identified White as best describing their racial background and stated "no" to being of Hispanic or Latino origin. Participants in the Latino group were those who identified as being of Hispanic or Latino origin, regardless of their Hispanic/Latino background (answer choices: Mexican/Mexican American, Cuban/Cuban American, Puerto Rican, Central/South American, Other Hispanic). For the exploratory aim, Latinos were categorized according to their identification with their Hispanic/Latino subgroup; Latinos who identified as being in more than one of these subgroups were excluded for these analyses. Our final sample included 9,919 Whites and 3,208 Latinos.

Participants reported their age and gender. For socioeconomic status, we evaluated parents' level of education (highest reported between mother and father; parent-reported when available, and otherwise adolescent-reported).

Acculturation. There were no questions directly assessing acculturation, so we coded the following proxy measures (for Latinos only): language spoken at home (English = 1, Other = 0), language, country of birth (U.S. = 1, Other = 0), parents' country of birth (U.S. = 1, Other = 0), and length of stay in the U.S. (5 years or greater = 1, fewer than 5 years = 0; Greenman & Xie, 2008). Scores for each of these measures were calculated into percentage of maximum possible (POMP) scores, and greater values indicate greater U.S. acculturation.

CES-D. The CES-D is a 20-item measure developed by the National Institute of Mental Health's Center for Epidemiologic Studies (Radloff, 1977). Individuals rated how often they felt each of the symptoms in the past week on a scale of 0 to 3 (e.g., 0 – *Rarely or none of the time, less than one day*; 3 – *most or all of the time, 5-7 days*). The positive affect items are reverse coded, and scores are summed, so they range from 0 to 60. Higher scores indicate greater

symptom severity. The cut-off score of 16 has been used to indicate individuals with a high risk of having depression (Radloff, 1977; Zich, Attkisson, & Greenfield, 1990). The CES-D has shown to have high reliability, with coefficients ranging from 0.85 to 0.90 (Radloff, 1977). Add Health used a modified version of the CES-D. The wording was changed from "I" to "you", four of the items (Items 7, 11, 17, 20) were changed to be more meaningful to adolescents (see Appendix 1 for summary of changes), and the "trouble staying asleep or falling asleep" and "frequent crying" were for the past 12 months and scored from 0-4 (e.g., 0 – *never;* 4 – *every day*). To account for these changes, we recoded these items so that responses of 3 or 4 will be marked as a 3 to make them consistent with the other 18 items. Despite the potential concern for these two items with the different time frame to form their own factor, previous research did not find a substantial difference in factor loadings from those items with the 1-week time frame (Crockett et al., 2005).

Analytic Plan

Statistical analyses followed recommendations outlined in Bauer (2017) and Curran et al. (2014). We used M*plus* Version 8 (Muthén & Muthén, 2007) and the *aMNLFA R* package for the for the EFA and MNFLA analyses (Cole et al., 2018; Gottfredson et al., 2018). For the measurement invariance analyses, *aMNFLA* produced individual output files, which were run using the *MplusAutomation R* package (Hallquist & Wiley, 2018). First, we examined the descriptive statistics and conducted graphical analyses of the 20 CES-D items and take note of any trends occurring as a function of each covariate of interest (age, gender, parent education). Next, we used EFA to test dimensionality and item local dependence, focusing on whether the two-factor structure proposed by Edwards et al. (2010) is supported in this data.

For aim 1, we tested two MNLFA models for each CES-D factor, allowing the factor means, intercepts, and factor loadings to vary by ethnicity (Latino vs. Whites). We did not test for variance impact. We compared the intercept and factor loadings for each item and determine whether any significant ethnicity DIF is present. For aims 2 and 3, we tested two additional MNLFA models in which we include ethnicity as well as other covariates (*gender, age, parent education, gender x ethnicity interaction*) and examine whether DIF occurs as a function of any of these covariates in the factor loadings or intercepts for the CES-D items.

To identify items showing DIF for each MNLFA model, we used an iterative strategy illustrated in Bauer (2017), with the assistance of the *aMNLFA* R package, in which first assumed all items to be invariant and then test for DIF with the covariates of interest (*ethnicity,* for the first model set, and *ethnicity, gender, age, parent education, and gender x ethnicity* interaction for the second model set). We assessed each CES-D item individually by freeing its parameters, while holding the other items constant as anchors. Once the CES-D item for which DIF would result in the largest improvement of fit was identified, we evaluated whether allowing DIF in a second item would significantly improve model fit, and so forth until no further improvement in model fit was possible. Any nonsignificant DIF terms were trimmed (using Wald tests), and we used the Benjamini-Hochberg procedure to adjust the Type I error rate (Benjamini & Hochberg, 1995; Thissen, Steinberg, & Kuang, 2002).

We fit a final set of MNLFA models (for each aim) by incorporating the specifications we determined for the factor means, variances, and DIF, such that all the significant effects will be incorporated into one MNLFA model for each factor (negative and positive). Lastly, we calculated factor scores for each individual, based on the parameters from each set of MNLFA models, and we compared the criterion validity of these scores by computing correlations with

three variables that may be related to depressive symptom severity: suicidal ideation endorsement (*During the past 12 months, did you ever seriously think about committing suicide?*), missing school (*In the last month, how often did a health or emotional problem cause you to miss a day of school?*), and receipt of counseling (*In the past year, have you received psychological or emotional counseling?*).

To assess the effect of DIF on the items and the overall measure, we visually inspected the following plots: item characteristic curves (ICCs) for items that showed the largest DIF by ethnicity (for each model) and test information curves, which is the amount of information that all the items in sum provide at any level of the latent trait. For the exploratory aim, we categorized Latinos into subgroups based on how they identified (Mexicans, Puerto Ricans, Cubans) and tested two MNLFA models examining whether DIF exists as a function of race, age, gender, parent education, or acculturation.

RESULTS

Demographics and Preliminary Analyses

Respondents in our sample (N = 13,127) ranged from 11 to 21 years of age (M = 15.7). The breakdown between males and females was almost evenly split (6661 females, or 51%). Table 1 presents the demographic characteristics between Latino and non-Latino White respondents. Latino respondents were significantly older and have higher CES-D total scores (all ps < .001). White respondents were significantly more likely than Latino respondents to have a parent with a college degree (26% compared to 13%, respectively χ^2 [1] = 235.54, p < .001). Non-Latino Whites were significantly more likely to report receiving counseling for psychological or emotional problems (14% versus 12%, respectively; p = .003). There were group differences in days missed from school in the past week due to physical or emotional

26

problems. Examination of the adjusted standardized residuals indicated that Latinos who endorsed missing a few times through almost every day and Whites who endorsed never missing school are observed significantly *more* than expected ($ps < .05$). In addition, Latinos who endorsed never missing school and Whites who endorsed missing a few times through almost every day are observed significantly less than expected ($ps < .05$). There were no group differences in gender or suicidal ideation ($ps > .47$).

MNLFA Testing

The distribution of the 20 CES-D items (scored 0 - 3) were assessed to ensure no cells were sparse and that the skew and kurtosis for each item was under reasonable limits (skew < 2, kurtosis < 7). All items met these criteria, with the exception of item 9 (*failure*), which had a skew value of 2.97 and kurtosis of 9.64. Therefore, we opted not to dichotomize the items and kept them as ordinal for the subsequent analyses. Our EFA results supported the two-factor structure proposed by Edwards et al. (2010), in which the 16 negatively worded symptoms loaded on one factor and the four positively worded items (reverse-coded) loaded on a second factor. This factor structure was consistent when looking at both the pooled data and also by ethnic subgroup (Latinos and Whites). Therefore, we conducted separate MNLFA models for each factor.

Model 1 (ethnicity only). Table 2 shows the complete set of estimated effects for the factor means, item intercepts, and factor loadings that showed in model proposed in Aim 1 (only examining ethnicity). The significant effects of ethnicity on the factor means indicate that Latinos reported more severe depressive symptoms than White respondents ($ps < .001$). None of the items in the positive factor showed any ethnicity DIF, whereas ethnicity intercept DIF was present for eight of the sixteen negative factor items (*bothered, blues, failure, sleep, talk less,*

unfriendly, crying spells, and *get going*). At similar levels of depression, Latinos were more likely to score higher on being bothered by things, feeling unable to shake the blues, talking less, and feeling like a failure. However, they were less likely to endorse having sleep problems, feeling like others were unfriendly, having crying spells, or difficulty getting started on things. The only items that showed ethnicity loading DIF were the *blues* and *crying spells* items; the factor loading for *blues* was lower for Latino respondents, so for Latinos, feeling like they couldn't shake the blues is a less core feature to the negative factor. For the *crying* item, the factor loading was greater in Latinos, suggesting that the item was more discriminating for Latinos than for Whites. In contrast with our Aim 1 hypothesis, Latinos were *less* likely to endorse two somatic items (*sleep*, *talk less*), and showed no differences for three items (*appetite*, *tired*, *mind*). There were also no group differences in the factor loadings for the somatic symptom items.

Model 2 (all covariates). Table 3 shows the complete set of estimated effects for the factor means, item intercepts, and factor loadings that showed in model proposed in Aim 2 (incorporating all covariates of interest). The factor means significantly differed by ethnicity, age, gender, and parental education (*p*s < .001). When comparing the ethnicity DIF from the previous model, the ethnicity intercept DIF from the *bothered* and *crying spells* items was not present in Model 2; instead, these items displayed both age and gender intercept DIF, which suggests that these characteristics may better explain the differences in responses.

Two items in the negative factor showed no differences across ethnicity, age, gender, parental education (*depressed* and *tired*). Items in the positive factor did not show any ethnic differences, whereas six items in the negative factor (*blues, failure, sleep, talk less, unfriendly, get going)* showed ethnic differences in the intercept loadings. At similar levels of depression,

Latino respondents were more likely to endorse feeling it was hard to shake off the blues, feeling like a failure, and talking less. They were *less* likely to endorse trouble sleeping, feeling people were unfriendly, and getting started on doing things (as indicated by negative intercept values), as compared to White respondents. With one exception (*blues),* there were no ethnic differences in the factor loadings for any of the items in the positive or negative factors, suggesting that the strength of the relationship between these items and their respective factor did not differ between Latino and White respondents. For the *blues* item, the factor loading on the negative factor was lower for Latinos, suggesting that feeling like it was hard to shake the blues is a less central feature (and less discriminating) within the negative factor for depression among Latino respondents as compared to White respondents.

Gender DIF was present for ten items in the negative factor (*bothered, appetite, mind, failure, talk less, unfriendly, crying spells, sad, dislike me, get going*), and one item in the positive factor (happy). For the most part, this gender DIF was confined to the intercepts, with the exception of gender loading DIF in the *crying* and *dislike* items in the negative factor and *happy* in the positive factor. There was no significant gender by ethnicity interaction in any of the items, so our Aim 3 hypothesis was not supported. At similar levels of depression, female respondents were more likely to endorse being bothered by things, having a poor appetite, crying frequently, and feeling sad. In contrast, the negative intercept loadings indicated that across similar levels of depression, female respondents reported lower severity scores for trouble keeping their mind on what they were doing, feeling like a failure, talking less, feeling people were unfriendly, feeling like people disliked them, and getting started doing things.

Age DIF was observed in the intercepts for nine items in the negative factor (*bothered, appetite, fearful, sleep, lonely, crying, dislike, get going)* and one positive factor item (*enjoyed*

life). At similar levels of depression, older respondents were more likely to endorse being bothered by things, having a loss of appetite, feeling lonely, and getting started doing things. Older respondents were less likely to endorse enjoying life. In addition, younger respondents were more likely to endorse being fearful, having sleep problems, and feeling that others dislike them. Age factor loading DIF was also observed for the *unfriendly* and *enjoyed life* items.

Lastly, parental education (high school and/or college education) intercept DIF was observed for nine items in the negative factor (*appetite, failure, sleep, lonely, unfriendly, dislike me, get going*) and two items in the negative factor (*good, hopeful*). More specifically, at similar levels of depression, respondents with a parent who has a college degree (or beyond) were less likely to endorse loss of appetite, feeling like a failure, or feeling like others dislike them; they were more likely to endorse having sleep problems, feeling lonely, feeling others were unfriendly, and trouble getting started on things. Respondents who have a parent with no high school degree were more likely to endorse feeling like a failure, feeling good, and feeling hopeful, while also being less likely to endorse having sleep problems, feeling lonely, and feeling that others were unfriendly. The *unfriendly* and *dislike* items also showed factor loading DIF with parental education: for respondents who have a parent with a high school degree (and above), feeling that others were unfriendly was a less discriminating item, whereas for respondents who have parents with at least a college degree, feeling that others disliked them was more discriminating across latent levels of the negative factor.

DIF Impact – Graphical Analyses

Figures 1 and 2 show the ICCs for select items in both model sets that showed the highest degree of DIF. In both models (Figures 1a and 1b, as well as 2a and 2b), the trace lines for *blues* and *failure* items are shifted slightly to the left for Latinos (dashed line) at each response

category. Figure 1c shows the trace lines for the *crying* item, which shows that the lines for

Latinos (dashed) were shifted slightly to the right, indicating that at similar levels of the negative

factor, Latinos were less likely to endorse crying frequently; however, this effect was no longer

significant in Model 2. Figure 2c shows the trace lines for the *unfriendly* item, which were

shifted slightly to the right for Latinos, indicating that they had to have levels of the negative

factor for them to report that people were unfriendly to them.

Figures 3a through 3d shows the total information curves for Models 1 and 2 (by factor),

which illustrate how well the test is doing at estimating the latent trait (negative or positive

factor) over the range of CES-D factor scores. In all four quadrants, the curves for Latinos and

Whites almost perfectly overlap, suggesting that the items characterize the negative and positive

latent trait similarly for Latino and White youths in both MNLFA models. Taken together, these

plots indicate that even though intercept and loading DIF was observed for various items, the

effects of DIF appears to be small and unlikely to be clinically meaningful when comparing

Latinos with White respondents.

Factor Score Comparisons

For each MNLFA model set, we computed factor scores and compared them with the

traditional unweighted scoring (sum of all the item scores, positive factor items being reverse

scored). Table 4 shows the correlations between the various scoring methods with endorsement

of suicidal ideation, days of school missed, and receipt of psychological counseling. For the

negative factor items, the traditional scoring approach (summing all the negative items) showed

the highest correlations for all three criterion variables, and the MNLFA factor score correlations

were similar; the Model 1 factor score correlation with suicidal ideation was higher than Model

2's ($p < .05$). For the positive factor, however, a different pattern emerged. Model 2 factor scores

had the highest correlations with all three variables ($ps < .05$). The traditional scoring method (all items, summed) was included in the table to illustrate the correlations between scores typically used in clinical and research settings with these outcome variables.

Exploratory Analyses

For our exploratory aim, we looked at respondents who identified as either Mexican ($n=1,514$), Cuban ($n = 457$), or Puerto Rican ($n = 508$). Latinos who identified as Central American ($n = 296$) or other Latino ($n = 205$) were excluded from the MNLFA analysis due to the sparse number of responses for some of the CES-D item categories. In addition, Latinos who indicated more than one subgroup ($n=216$) were not included in these exploratory analyses. Table 5 shows the demographic characteristics (age, gender, acculturation variables, parental education, CES-D total scores) between the remaining groups. One way between-subjects ANOVAs were conducted to examine group differences in age, CES-D total scores, and acculturation POMP scores. There was a significant effect of group on age, $F(2, 2469) = 24.27$, p $< .001$. Games-Howell post hoc tests showed that Cubans were significant older than Mexicans and Puerto Ricans, and Puerto Ricans were significantly older than Mexicans (all $ps < .001$; see Table 4). There was also a significant effect of group on CES-D total scores, $F(2, 2469) = 12.27$, $p < .001$. Mexicans and Puerto Ricans both had higher CES-D scores than Cubans ($ps < .01$), but they did not differ significantly from each other ($p = .47$). Lastly, there was a significant group effect for acculturation POMP scores, $F(2, 2466) = 251.59$, $p < .001$. Post hoc tests indicated that Puerto Ricans had significantly higher acculturation scores, followed by Mexicans ($ps < .001$). Chi-square test indicated that parent education significantly differed by group, $\chi^2(4) = 130.79$, $p < .001$. Examination of the adjusted standardized residuals indicated that Mexicans who have a parent with no high school degree, Puerto Ricans who have a parent with a high school degree,

and Cubans who have a parent with a college degree (or above) are observed significantly more than expected, whereas Mexicans who have a parent with a high school degree or above, and Cubans and Puerto Ricans who have a parent with no high school degree are observed significantly less than expected (ps < .05).

Table 6 shows the factor means and DIF parameters between the three Latino subgroups (Mexicans, Cubans, and Puerto Ricans). For both negative and positive factors, the means differed significantly by race, age, and gender. For the positive factor, respondents with parents who had at least a high school education reported lower symptom severity (p = .002). Race intercept DIF was present for two negative factor items (*fearful* and *dislike*) and two positive factor items (*good* and *hopeful*). At similar levels of the negative factor, Cubans were less likely to endorse being fearful and feeling disliked by others. For the positive factor items, Cubans and Puerto Ricans were *more* likely to report feeling good, and Cubans were more likely to report feeling hopeful, relative to the other Latino groups. Acculturation intercept DIF was present for three negative factor items (*bothered, tired, sleep,* and *dislike me*). However, the intercepts only differed by 0.01 units, so the difference was small. Five items in the negative factor (*blues, depressed, failure, lonely,* and *sad*) and all but one item in the positive factor (*good, happy, enjoyed life*) showed no differences across race, age, gender, parental education, and acculturation within the Latino sample. Gender intercept DIF was present for six negative factor items (*appetite*, *mind*, *talk less*, *unfriendly*, *crying spells*, and *get going*) but for none of the positive factor items. Lastly, there was age intercept and loading DIF for the *good* item in the positive factor, suggesting that for older Latino respondents, feeling good is a less core feature of the positive factor than for younger Latino respondents.

DISCUSSION

Although many studies have explored culture DIF in various depression scales, including the CES-D, there has not been a consensus on symptoms that consistently show DIF between Latinos and non-Latino Whites. There are many possible explanations for this, including variation in how (if at all) other demographic variables were accounted for in the analyses, as well as statistical technique used to test for DIF, with its own assumptions, strengths, and limitations. This variation makes comparing results across studies challenging, especially if the background characteristics of the samples are vastly different outside of the target characteristic (i.e., age, gender). Therefore, this investigation aimed to use MNLFA to examine the role that culture plays in depressive symptom presentation between Latino and White youths on the CES-D, while incorporating other demographic covariates that may also explain differences in endorsement of depressive symptoms. To the best of our knowledge, this was the first study to use MNLFA to closely examine patterns in depressive symptom endorsement between Latino and non-Latino White youths, while also paying attention to how other demographic characteristics influence symptom presentation between these groups.

Main Findings

For our first aim, we examined whether there were cultural differences in the endorsement of somatic symptoms to help elucidate previous mixed findings on patterns in somatic symptom endorsement (e.g., Choi & Park, 2006; Kirmayer, 2001; Lewis-Fernández et al., 2005). We hypothesized that Latinos will be more likely to endorse somatic symptoms (at similar levels of depression) and show higher factor loadings for these items than Whites. However, in both sets of MNLFA models, the only somatic item Latinos were more likely to endorse was talking less; otherwise, they were *less* likely to endorse having sleep problems and

34

trouble getting started on activities, and they showed no differences in endorsement of feeling tired, keeping their mind on what they were doing, or appetite loss. In addition, the factor loadings for these somatic items did not differ between Latinos and Whites, indicating no cultural differences in discrimination between low and high levels of the underlying latent trait. These findings support previous work that also failed to find differences in the rates of somatic symptom endorsement (Kirmayer et al., 1993; Kroenke et al., 1997; Simon et al., 1999; Uebelacker, Strong, Weinstock, & Miller, 2009). Kirmayer (2001) argued that setting plays an important role on what symptoms individuals are likely to report on initially; in a primary care office, for example, respondents are likely to focus on physical/somatic symptoms even when they are depressed potentially because of beliefs about the relevance of the more internalizing/emotional symptoms. Therefore, initial somatic complaints does not necessarily indicate that those symptoms are the individuals' sole or primary depressive symptoms (Kirmayer, 2001; Kirmayer et al., 1993; Uebelacker et al., 2009). This argument is in line with what we observed; the Add Health respondents completed the CES-D interview questions in their homes, were not treatment-seeking individuals, and they answered standardized questions with specific response options.

Our second aim was to examine whether items still showed ethnicity DIF after accounting for other important demographic characteristics, such as age, gender, and parental education. We hypothesized that a MNLFA model that incorporates these demographic covariates would provide a more accurate picture of any cultural differences than a model that only looked at ethnicity DIF because DIF for some items may be better explained by other demographic variables. When comparing the two MNLFA model sets, two items that had initially showed ethnicity DIF in the negative factor (being bothered by things, crying spells), no

longer showed ethnicity DIF in the full model. Instead, these items showed both age and gender intercept DIF in the intercepts, indicating that age and gender led to differences in endorsing these items at similar levels of depression. Latinos in this sample were significantly older than the White respondents, so one possibility is that not incorporating age in the first model may have conflated the source of DIF for these items. This illustrates one of the strengths of MNLFA; by looking at multiple covariates simultaneously when assessing DIF, one can avoid potentially confounding the individual effects of one characteristic that may be driven by another, which can cloud our understanding of any cultural differences. In addition, several items not only showed ethnicity DIF, but DIF for gender, age, and parental education, indicating that DIF coming from various characteristics can be occurring simultaneously.

One effect we observed in both models was the presence of intercept and factor loading DIF for Latinos with the *blues* item. In particular, the factor loading was lower by 0.35 units for Latinos than for Whites, which indicates that feeling unable able to "shake off the blues, even with the help from family and friends" is not as discriminating between Latinos with lower and higher levels of the negative latent trait, as compared to White respondents. One possible explanation for DIF being present in this item is because of the way it is written: "shake off the blues" is an idiom, so some individuals, especially those who did not grow up speaking English at home or were raised in another culture, may not be familiar that *the blues* means feeling sad or depressed, so even if they feel severely depressed, they may not rate this item highly. When using an existing measure with a different culture, it is recommended that the questions be evaluated to ensure that they have idiomatic equivalence across different groups even if keeping the written language the same (Beaton et al., 2000). One way to investigate this hypothesis in a

future study is to re-word this question to remove any idioms to see if the culture DIF still remains.

When comparing the MNLFA factor scores with the traditional CES-D scoring by computing correlations with three established variables thought to be related to depressive symptomatology, our findings were mixed and not quite what we expected. For the positive factor, the Model 2 factor scores had the highest correlations with all three variables (suicidal ideation, missed school, receipt of counseling), and the traditional scoring had the lowest correlation, which is consistent with previous research that incorporating background characteristics improves the score quality (Curran et al., 2016). For the negative factor, however, the traditional score (summing all negative items) produced the highest correlations with the three outcome variables, and for the most part, the two model factor scores showed similar correlation coefficients. We did not expect that the traditional negative factor score would outperform the MNLFA factor scores, given that traditional scoring methods face many limitations and do not incorporate individual-level characteristics of the respondent. Although we can only speculate, one possibility is that the CES-D negative factor items include depression-irrelevant variance that is also related to the criterion variables we examined. For future analyses, we will assess how the negative and positive factors differ and whether estimating a multidimensional MNLFA model could lead to more accurate factor scores.

We also examined whether there were any interactions with gender and ethnicity on DIF with the CES-D items. We predicted that because Latino culture has placed an emphasis on particular gender norms, known as *Machismo* and *Marianismo*, we would find evidence of an interaction such that Latino boys would be *more* likely to endorse somatic symptoms than Latino girls, and that this gender difference would be less pronounced with White youths. This

interaction was not significant in any of our final trimmed models. However, when looking at

our exploratory model (only including Latinos), Latino males were *more* likely to endorse

difficulty concentrating, talking less, and getting started on activities, while Latina females were

more likely to report loss of appetite and crying spells. These gender differences were also

present when looking at the pooled sample. The largest effect was for the crying item, and this

gender DIF pattern has been well-established in the literature (e.g., Steinberg & Thissen, 2006;

Van Beek et al., 2012). These findings suggest that we do observe gender differences in

symptom endorsement, but these differences manifest similarly in the Latino respondents as

well. One thing to consider is that Add Health did not ask respondents about their *beliefs* about

gender norms, so an extension of this work could be to assess whether the patterns of DIF would

change when comparing an individual's beliefs versus their gender identity.

As an exploratory aim, we capitalized on a strength of the Add Health study and

examined DIF by Latino subgroup (Mexicans, Cubans, and Puerto Ricans), which has not often

possible in previous research due to various factors, including insufficient sample sizes and/or

researchers not collecting information about Latinos' racial identification (Alegría et al., 2008;

Cauce & Domenech-Rodriguez, 2002). To account for cultural differences within the Latino

subgroups, we created a composite acculturation measure based on multiple variables (language

spoken at home, self and parents' country of birth, years in the US) to operationalize degree of

acculturation. We hypothesized that less acculturated Latinos would endorse higher somatic

symptoms and be *more* likely to endorse positively worded CES-D items. However, we found no

connection between acculturation scores and endorsement of positive factor items. Acculturation

DIF seemed to appear in four negative factor items: being bothered by things, feeling tired,

having sleep problems, and feeling that others dislike them, with only two of these considered

somatic symptoms (sleep problems, feeling tired). The difference in the intercepts between Latinos who had higher acculturation scores was only a maximum of 0.01 units, so this difference is not likely to be clinically meaningful. We observed subgroup DIF with other items; at similar levels of depression, Cubans were *less* likely to endorse feeling fearful and being disliked by others, while being more likely to endorse feeling good and feeling hopeful than Puerto Ricans and Mexicans. Puerto Ricans were also more likely to endorse feeling good. There were notable differences in parental education and other demographic characteristics across the Latino subgroups, which we would have overlooked if we had only assessed Latinos as one group. Given these demographic differences, incorporating them into a MNLFA model allowed us to examine items that still showed culture DIF even after accounting for these variables, which helps us to better understand cultural differences in symptom presentation.

Strengths and Implications

One of the strengths of this study was having a large sample of Whites and Latinos, providing us with the statistical power to detect even small differences. The DIF findings that emerged remained statistically significant even after using the Benjamini-Hochberg procedure to control the false discovery rate. However, given our increased statistical power, we must critically evaluate the patterns to determine whether the DIF we observed in the CES-D items is clinically meaningful and warrants any changes to the measure format or scoring. Upon closer examining graphical depictions of item response functions and test information curves, we determined that these differences are not likely to have a large clinical impact on the CES-D scores. One possibility for why this occurs is that DIF at the item level canceled out at the overall score level because the direction of DIF between Latinos and Whites did not appear consistent in one direction; therefore, simply removing items that showed DIF, such as the *blues* item that

showed consistent DIF in both MNLFA models, could have negative consequences and compromise the scale validity (Mchorney & Fleishman, 2010). When comparing the mean factor scores with the traditional unweighted scoring, both sets of scoring indicated that Latinos in this sample had higher mean CES-D scores, indicating that even after accounting for DIF, Latino youths in this sample reported slightly higher depressive symptom severity. The direction of the results did not change, indicating that the CES-D scoring may not need to be adjusted. However, given that this study was a community sample rather than with a group of youths seeking treatment for mental health problems, the patterns of DIF may change with youths who have more severe psychopathology or have CES-D scores on the higher end of the spectrum.

Another unique contribution our study has added to the literature is the examination parental education as a proxy for SES in depression symptom endorsement. By incorporating parental education in our MNLFA models, we found that respondents who had parents with at least a high school degree had lower negative and positive factor (reverse-coded) means (with college degree or more having the lowest depressive symptom severity), and when accounting for the other covariates, respondents with a parent with a high school degree (and above) and a college degree (or above) showed DIF in endorsement of appetite, feeling like a failure, sleep problems, feeling lonely, feeling that others were unfriendly, trouble getting going, as well as feeling hopeful and good. The direction of the DIF did not go in one consistent direction, so it is likely that the effects on the overall measure is small and likely to cancel out. However, it is nonetheless important to consider why we might be seeing these patterns. Some researchers have argued that groups with less power (as defined by lower SES) may be more attune to impression management and thus be less likely to report on any stigmatizing symptoms (Johnson & Van de Vijver, 2002; Ross & Mirowsky, 2006). Our DIF findings did not provide clear evidence for this

argument; for example, for the *failure* item, respondents who had a parent with no high school education were *more* likely to endorse feeling like a failure, but relative to the other groups, they were *less* likely to endorse problems with sleep, feeling that others were unfriendly, as well as less likely to endorse feeling good and hopeful. Furthermore, respondents with a parent who had no high school education also had higher CES-D mean scores for both the positive and negative factors, indicating greater depressive symptom severity. This is in line with our argument that having a parent with less education may indicate being of a lower SES, which in turn is associated with increased stressors that can exacerbate mental health symptoms (Conger et al., 1994; Goodman, Meltzer, & Bailey, 1998). Observing the degree of items that showed education DIF illustrates the need to more closely examine the question of whether impression management in mental health measures differs as a function of SES.

Limitations

These study findings, however, need to be considered in the context of several limitations. First, we did not use sample weights, which is important given that several groups of adolescents were selected with unequal probability based on race (e.g., Cubans, Puerto Ricans), as well enrollment in certain schools. Therefore, estimates will be biased and one cannot draw conclusions about the general population (Chen & Chantala, 2013). For these analyses, we ultimately decided not to use the sample weights because we were only looking at a subset of the Add Health sample (i.e., Whites and Latinos), and because the sampling weights assume a whole-sample analysis, we would have to re-norm the weights. Furthermore, no previous study has used weighting when conducting MNLFA, so further work needs to be done to ensure the correct weighting procedures are incorporated to be able to draw conclusions about the population. In addition, we opted not to test for variance impact in our models, so we were not

able to comment on how the variances of the factors change depending on the covariates in our models. We ultimately decided not to test for this because we did not have set hypotheses about how our covariates might influence the variance, but we may decide to include it in future model iterations, especially since we observed factor mean differences.

The available variables we used to operationalize SES and acculturation were proxies, so they may not fully capture the nuances of these constructs. For SES, we had originally wanted to use parent-reported family income and receipt of public assistance (e.g., food stamps, housing subsidy, or Aid for Families with Dependent Children [AFDC]) in addition to parent education, but due to the degree of missing data (20-30% for family income, 11-18% for receipt of financial assistance) and to reduce complexity of the MNLFA models, we ultimately decided to use parent education as the sole proxy for SES. Researchers argue that SES is a multidimensional construct comprising several factors, which include individual, household, and/or neighborhood characteristics, but unfortunately, there is no clear consensus on the best measures to examine it (Braveman et al., 2005; Shavers, 2007). However, our rationale for using parental education as a proxy for SES was because higher levels of education tend to be associated with more economic resources (APA, 2007), and more education can be indicative of better understanding of mental health symptoms, which in turn, could mean increased likelihood of seeking treatment. Previous studies using Add Health have also used parental education as a measure for SES due to the missing data in family income (Stewart & Reed, 2015). However, although education and income tend to be highly correlated, they still contribute different information and thus are not interchangeable (Braveman et al., 2005). Therefore, future iterations of these analyses should explore creating composite scores with multiple indicators of SES to see which DIF patterns observed with parental education would stay consistent.

Regarding acculturation, an ideal measure would have been to assess Latino respondents' degree of alignment or identity with their home culture as well as US culture (Cabassa, 2003). However, Add Health did not use any formal acculturation measure, so we created a composite score from proxies for acculturation (i.e., country of birth, length in the US, language spoken at home), which have been used in previous studies (Alegria, 2009; Greenman & Xie, 2008). By using multiple items to assess for acculturation, we have the added advantage of having more content coverage, rather than relying on one item to define a complex construct.

In our main MNLFA analyses, we included biracial participants if they responded that White was the "category that best described [their] racial background" or if they identified as Hispanic/Latino. 311 White respondents and 215 Latino respondents indicated more than one race, but we decided to include them in the analyses to not only make the findings more generalizable, but because many Latinos may identify as mixed-race due to Latin America's colonial history that can include having White European, African, and Native American roots (Gonzalez-Barrera, 2015). There is a lack of research examining DIF among biracial individuals, so it is unclear how each identity may independently influence (if at all) an individuals' responses to self-report measures, so one natural extension is to run subsequent MNLFA models only including respondents who reported one race to see how the pattern of DIF may change.

Conclusions

Although many of our hypotheses were not supported, using MNLFA to examine culture DIF provides many fruitful lines of inquiry for future research. First, sample weights should be incorporated into the MNLFA models to ensure we can extend any conclusions about the general population. Furthermore, an additional step recommended by Bauer (2017) when you have a multidimensional measure is to fit a final MNLFA model combining the factors. Due to

computational burden and our primary focus being to critically evaluate DIF, we opted to look at the factors in separate models, while acknowledging that we could not test for the covariance between the factors or have an overall CES-D multifactor score. Therefore, a natural extension will be to combine our trimmed models and examine whether having one overall MNLFA score performs better than the traditional CES-D scoring.

Using MNLFA to examine DIF in the CES-D provided a more comprehensive picture on how various demographic characteristics can influence an individual's response to a depression questionnaire; although still computationally challenging to estimate these models, the creation of the *aMNLFA R* package has made it easier for researchers to implement this technique to explore similar research questions (Cole et al., 2018). The DIF we observed with the CES-D did not appear to be large, which suggests that the CES-D does not have to be changed in order to function similarly with Latinos.

Table 1

Descriptive statistics for demographics by ethnic group

	Non-Latino Whites ($n = 9,919$)	Latinos ($n = 3,208$)	Effect Size
	N (%) or *M* (*SD*)	*N* (%) or *M* (*SD*)	
Age	15.6 (1.73)	16.0*** (1.70)	$d = 0.24$
Female	5,051 (51%)	1,610 (50%)	$\phi = -0.01$
CES-D Total Score	11.8 (7.86)	13.8*** (8.31)	$d = 0.26$
Parent Education Level			$\tau = -0.27$
Did not graduate high school	952 (10%)	1,372 (43%)	---
Completed high school or trade school	6,295 (64%)	1391 (44%)	---
College degree or higher	2,562 (26%)	411 (13%)	---
Suicidal Ideation (12 months)	1358 (14%)	433 (14%)	$\phi = 0$
Missed School Days (1 month)			$\tau = 0.04$
Never	6580 (66%)	1941 (61%)	---
Just a few times	2869 (29%)	1077 (34%)	---
About once a week	319 (3%)	136 (4%)	---
Almost every day	59 (0.6%)	30 (0.9%)	---
Every day	48 (0.5%)	13 (0.4%)	---
Receipt of Counseling (12 months)	1380** (14%)	381 (12%)	$\phi = -0.03$

Note: Where data points were missing, means and percentages were calculated out of total number of available cases. Because we pro-rated CES-D scores if someone skipped one or two items, it is possible to have a non-integer total score.

**$p<.005$, ***$p < .001$, two-tailed, based on Welch's *t*-test (means) or X^2 (proportions).

Table 2

Final trimmed MNLFA model with significant DIF for intercept or factor loadings – Ethnicity only (Model 1)

Item *Covariate Effect*	Intercept (*SE*)	Loading (*SE*)	Item *Covariate Effect*	Intercept (*SE*)	Loading (*SE*)
Negative Factor					
1. Bothered	0	1.52 (0.03)	11. Sleep	0	0.72 (0.02)
Latino	0.06 (0.05)	---	*Latino*	-0.40 (0.04)	---
2. Appetite	0	1.09 (0.03)	13. Talk less	0	0.88 (0.02)
3. Blues	0	2.59 (0.06)	*Latino*	0.31 (0.05)	
Latino	0.28 (0.08)	-0.35 (0.08)	14. Lonely	0	2.14 (0.04)
5. Mind	0	1.25 (0.03)	15. Unfriendly	0	0.91 (0.03)
6. Depressed	0	3.00 (0.06)	*Latino*	-0.28 (0.05)	---
7. Tired	0	1.11 (0.02)	17. Crying Spells	0	1.16 (0.03)
9. Failure	0	1.77 (0.04)	*Latino*	-0.39 (0.06)	0.20 (0.06)
Latino	0.58 (0.06)	---	18. Sad	0	2.58 (0.05)
10. Fearful	0	1.36 (0.03)	19. Dislike Me	0	1.26 (0.03)
			20. Get "Going"	0	0.98 (0.02)
			Latino	-0.22 (0.04)	---

Factor Mean:	**Estimate (*SE*)**	***p***			
Latino	0.23 (0.02)	< .001			

Positive Factor (Reverse Coded)					
4. Good	0	1.25 (0.03)	12. Happy	0	2.32 (0.05)
8. Hopeful	0	1.33 (0.03)	16. Enjoyed Life	0	2.58 (0.06)

Factor Mean:	**Estimate (*SE*)**	***p***			
Latino	0.34 (0.02)	< .001			

Note. Only significant DIF parameters are tabled (based on Benjamini-Hochberg Procedure). Age is centered at 16 years old. Factor variance was fixed to 1 (not set to vary as a function of the covariates). Intercepts are set to 0. Positive factor items are reverse coded so higher scores indicated more severe depressive symptoms.

Table 3

Final trimmed MNLFA model with significant DIF for intercept or factor loadings – All covariates (Model 2)

Item / *Covariate Effect*	Intercept *(SE)*	Loading *(SE)*	Item / *Covariate Effect*	Intercept *(SE)*	Loading *(SE)*
Negative Factor					
1. Bothered	0	1.45 (0.03)	13. Talk less	0	0.93 (0.02)
Age	0.10 (0.01)	---	*Latino*	0.31 (0.04)	
Female	0.22 (0.04)	---	*Female*	-0.61 (0.04)	
2. Appetite	0	1.01 (0.03)	14. Lonely	0	2.08 (0.04)
Age	0.05 (0.01)	---	*Age*	0.08 (0.01)	
Female	0.52 (0.04)	---	*College*	0.28 (0.06)	
College	-0.17 (0.05)	---	15. Unfriendly	0	1.12 (0.06)
3. Blues	0	2.50 (0.06)	*Latino*	-0.28 (0.05)	---
Latino	0.30 (0.08)	-0.35 (0.08)	*Female*	-0.52 (0.04)	---
5. Mind	0	1.26 (0.03)	*Age*	-0.02 (0.01)[n.s.]	-0.07 (0.01)
Female	-0.37 (0.04)	---	*High school*	0.14 (0.07)	-0.22 (0.06)
6. Depressed	0	2.89 (0.06)	17. Crying Spells	0	0.98 (0.05)
7. Tired	0	1.08 (0.02)	*Age*	-0.08 (0.01)	---
9. Failure	0	1.75 (0.04)	*Female*	1.76 (0.05)	0.29 (0.06)
Latino	0.41 (0.07)	---	18. Sad	0	2.47 (0.05)
Female	-0.28 (0.06)	---	*Female*	0.25 (0.05)	---
High school	-0.46 (0.07)	---	19. Dislike Me	0	1.37 (0.05)
College	-0.24 (0.08)	---	*Age*	-0.12 (0.01)	---
10. Fearful	0	1.35 (0.03)	*Female*	-0.32 (0.05)	-0.18 (0.06)
Age	-0.06 (0.02)	0.06 (0.02)	*College*	-0.14 (0.06)	0.28 (0.07)
11. Sleep	0	0.72 (0.02)	20. Get "Going"	0	0.99 (0.02)
Latino	-0.27 (0.04)	---	*Latino*	-0.21 (0.04)	---
Age	-0.07 (0.01)	---	*Age*	0.04 (0.01)	---
High school	0.30 (0.05)	---	*Female*	-0.29 (0.04)	---
			College	0.20 (0.04)	---
Factor Mean:	Estimate *(SE)*	*p*		Estimate *(SE)*	*p*
Latino	0.12 (0.02)	< .001	*High school*	-0.16 (0.03)	< .001
Age	0.08 (0.01)	< .001	*College*	-0.11 (0.02)	< .001
Female	0.39 (0.02)	< .001			
Positive Factor (Reverse Coded)					
4. Good		1.22 (0.03)	12. Happy		2.14 (0.06)
High school	-0.27 (0.05)	---	*Female*	-0.64 (0.05)	0.41 (0.06)
College	-0.18 (0.05)	---	16. Enjoyed Life		2.52 (0.06)
8. Hopeful		1.28 (0.03)	*Age*	0.06 (0.02)	-0.08 (0.02)
High school	-0.29 (0.05)	---			
Factor Mean:	Estimate *(SE)*	*p*		Estimate *(SE)*	*p*
Latino	0.23 (0.03)	< .001	*High school*	-0.17 (0.03)	< .001
Age	0.03 (0.01)	< .001	*College degree*	-0.11 (0.02)	< .001
Female	0.23 (0.02)	< .001			

Note. Only significant DIF parameters are tabled (based on Benjamini-Hochberg Procedure). Age is centered at 16 years old. Factor variance was fixed to 1 (not set to vary as a function of the covariates). Intercepts are set to 0. Positive factor items are reverse coded so higher scores indicated more severe depressive symptoms.

Table 4

Correlations between CES-D scores and suicidal ideation, days missed from school, and receipt of counseling for emotional problems

Score Type	Suicidal Ideation	Missed School	Counseling
Model 1 Negative Factor	0.336[***]	0.247[***]	0.212[***]
Model 2 Negative Factor	0.335[***]	0.247[***]	0.212[***]
Traditional Scoring – Negative Factor	**0.353[***]**	**0.258[***]**	**0.230[***]**
Model 1 Positive Factor	0.248[***]	0.140[***]	0.124[***]
Model 2 Positive Factor	**0.253[***]**	**0.147[***]**	**0.127[***]**
Traditional Scoring – Positive Factor	0.245[***]	0.135[***]	0.114[***]
Traditional Scoring – All items	0.364[***]	0.251[***]	0.222[***]

***$p < .0001$. Bolded correlation indicates the highest out of category ($p < .05$).

Table 5

Descriptive statistics for demographics by Latino subgroup

|	Mexicans (M) (*n* = 1,514)	Cubans (C) (*n* = 457)	Puerto Ricans (P) (*n* = 508)	Effect Size		
	N (%) or *M* (*SD*)	*N* (%) or *M* (*SD*)	*N* (%) or *M* (*SD*)			
Age	16.0_a (1.74)	16.3_b (1.47)	15.6_c (1.67)	$\omega^2 = .02$		
CES-D Total Score	14.44_a (8.19)	12.22_b (8.86)	13.92_a (8.45)	$\omega^2 = .01$		
Acculturation – POMP Score	64.1 (31.7)	39.6 (25.7)	81.8 (23.5)	$\omega^2 = .17$		
Born in US	1218 (80%)	239 (52%)	458 (90%)	---		
English spoken at home	795 (53%)	109 (24%)	406 (80%)	---		
Mother born in US	618 (44%)	39 (9%)	332 (69%)	---		
Father born in US	412 (37%)	30 (10%)	193 (64%)	---		
Length of stay in US > 5 years	1432 (96%)	409 (90%)	494 (98%)	---		
				M-C	M-P	C-P
Female	746 (49%)	230 (50%)	253 (49%)	$\phi = 0.01$	$\phi = 0$	$\phi = -0.01$
Parent Education Level				$\tau = 0.12$	$\tau = 0.20$	$\tau = 0.08$
Did not graduate high school	814 (55%)	174 (38%)	139 (27%)	---		
Completed high school or trade school	530 (35%)	147 (32%)	211 (42%)	---		
College degree or higher	147 (10%)	41 (9%)	41 (8%)	---		

Note: If overall p-value from ANOVA < 0.05, Games-Howell post hoc tests examined group differences, so values with differing subscripts are significantly different at the *p* < .05 level. Omega squared, phi coefficients, and Stuart's Tau estimate effect sizes.

Table 6

CES-D differential item functioning (DIF) parameters from trimmed exploratory MNLFA model

| Item *Covariate Effect*	Intercept *(SE)*	Loading *(SE)*	Item *Covariate Effect*	Intercept *(SE)*	Loading *(SE)*
Negative Factor					
1. Bothered	0	1.45 (0.07)	13. Talk less	0	0.92 (0.05)
Acculturation	0.01 (0.001)	---	*Female*	-0.56 (0.09)	---
2. Appetite	0	1.00 (0.06)	14. Lonely	0	2.20 (0.10)
Female	0.43 (0.09)	---	15. Unfriendly	0	1.03 (0.06)
3. Blues	0	2.13 (0.10)	*Female*	-0.43 (0.10)	---
5. Mind	0	1.27 (0.06)	17. Crying Spells	0	0.84 (0.11)
Female	-0.34 (0.09)	---	*Female*	1.34 (0.12)	0.58 (0.13)
6. Depressed	0	2.79 (0.12)	18. Sad	0	2.52 (0.11)
7. Tired	0	1.15 (0.06)	19. Dislike Me	0	1.33 (0.07)
Acculturation	0.01 (0.001)		*Cuban*	-0.52 (0.14)	---
9. Failure	0	1.64 (0.08)	*Acculturation*	-0.01 (0.002)	---
10. Fearful	0	1.38 (0.07)	20. Get "Going"	0	1.07 (0.06)
Cuban	-0.39 (0.14)	---	*Female*	-0.44 (0.09)	---
11. Sleep	0	0.69 (0.05)			
Acculturation	0.01 (0.001)	---			
Factor Mean:	Estimate *(SE)*	*p*			
Cuban	-0.31 (0.06)	< .001			
Age	0.06 (0.01)	< .001			
Female	0.42 (0.05)	< .001			
Positive Factor (Reverse Coded)					
4. Good	0	1.04 (0.06)	8. Hopeful	0	0.98 (0.06)
Cuban	-0.41 (0.11)	---	*Cuban*	-0.32 (0.11)	---
Puerto Rican	-0.53 (0.10)	---	12. Happy	0	2.03 (0.12)
Age	-0.06 (0.03)	-0.08 (0.03)	16. Enjoyed Life	0	2.20 (0.13)
Factor Mean:	Estimate *(SE)*	*p*		Estimate *(SE)*	*p*
Cuban	-0.15 (0.07)	.03	*High school*	-0.16 (0.05)	.002
Age	0.05 (0.01)	.02	*Acculturation*	-0.001 (0.001)	.001
Female	0.24 (0.05)	.05			

Note. Only significant DIF parameters are tabled (based on Benjamini-Hochberg Procedure). Age is centered at 16 years old. Factor variance was fixed to 1 (not set to vary as a function of the covariates). Intercepts are set to 0. Positive factor items are reverse coded so higher scores indicated more severe depressive symptoms.
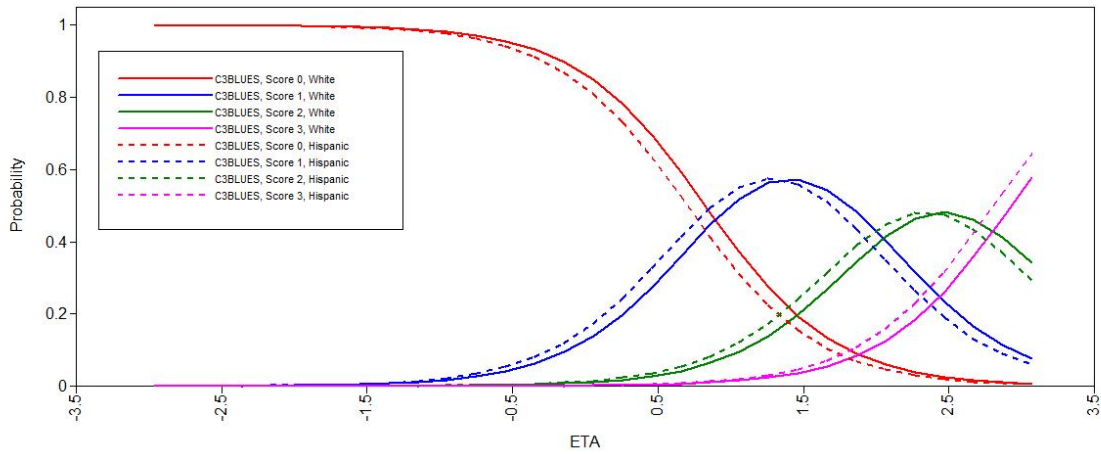
a) ICC for Item 3 – *You felt that you could not shake off the blues, even with the help from your family and friends.*
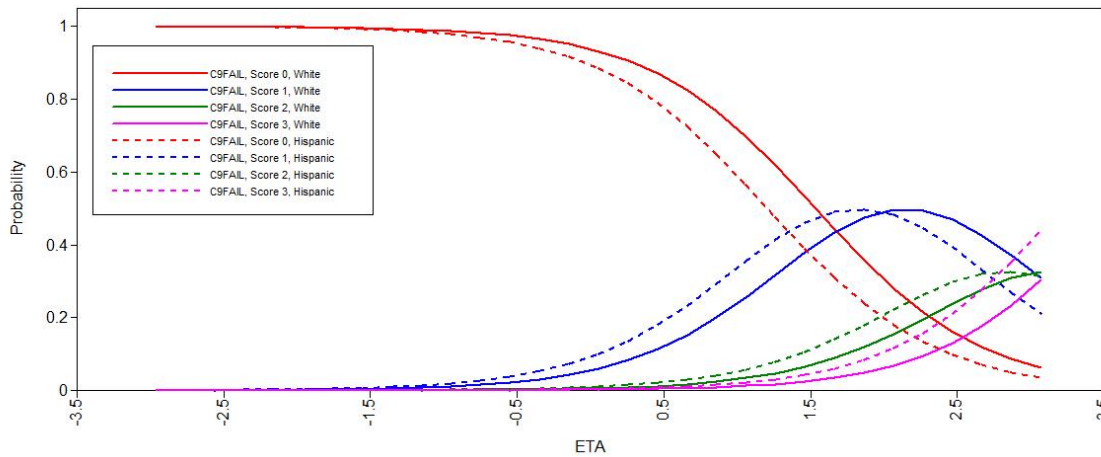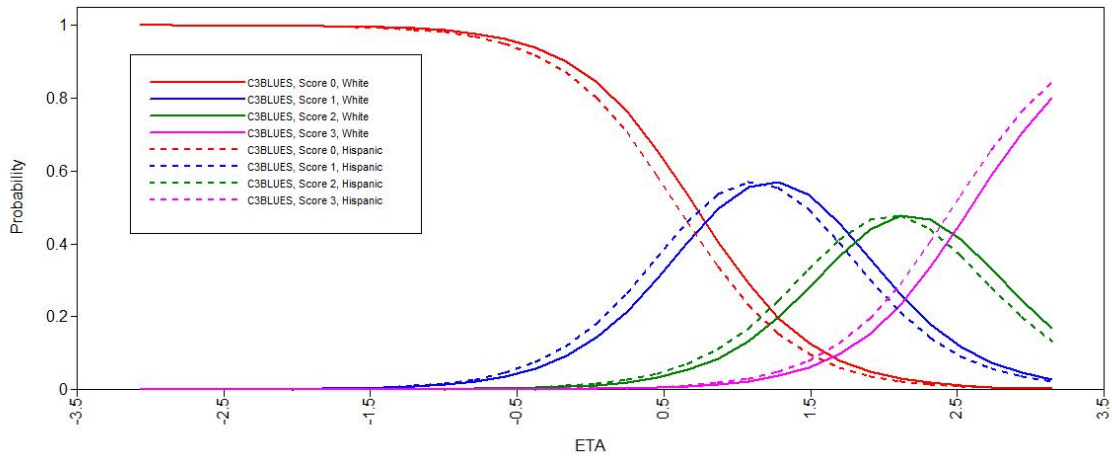

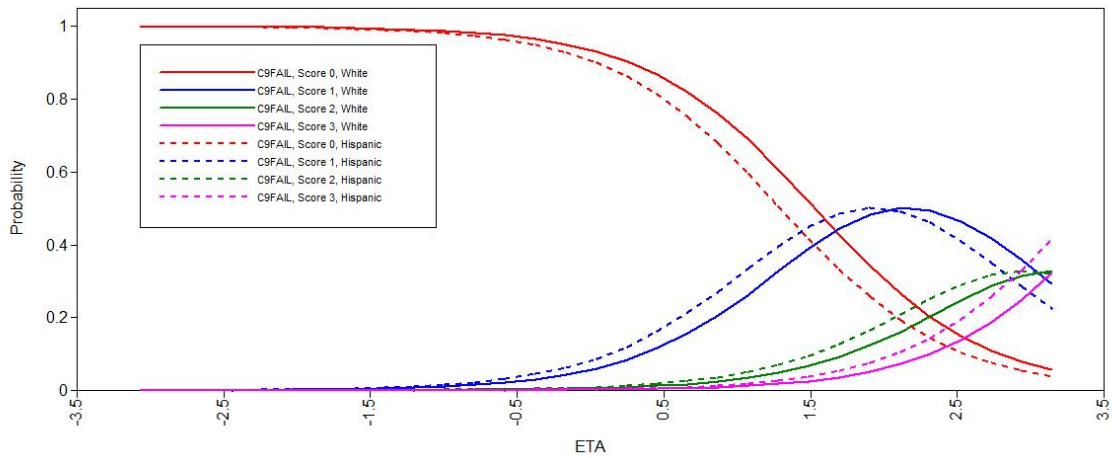b) ICC for Item 9 – *You thought your life had been a failure.*
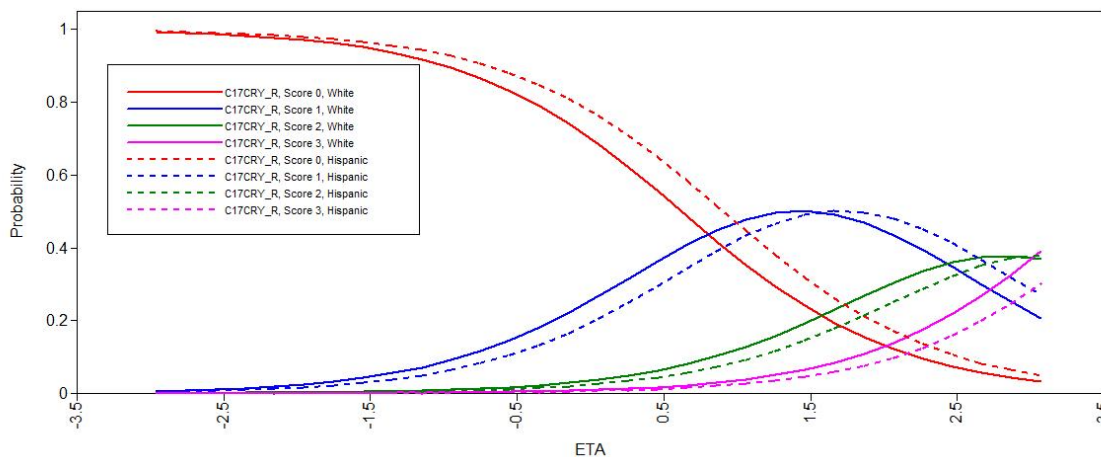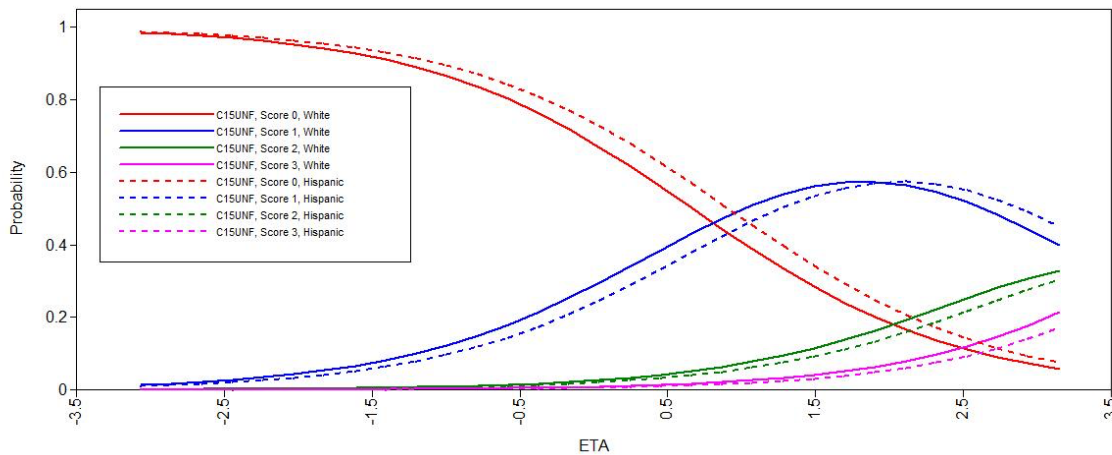

c) ICC for Item 17 – *You have cried frequently.*

Figure 1. Item Characteristic Curves (ICCs) for Latinos (dashed lines) and Whites (solid lines) for three items that showed largest DIF in Model 1.

a) ICC for Item 3 – *You felt that you could not shake off the blues, even with the help from your family and friends.*
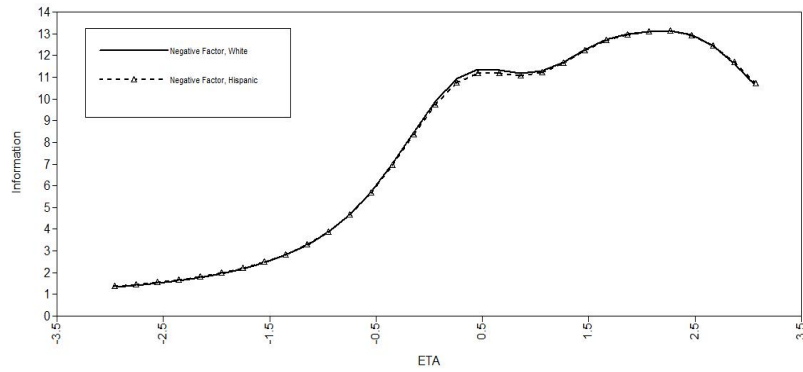

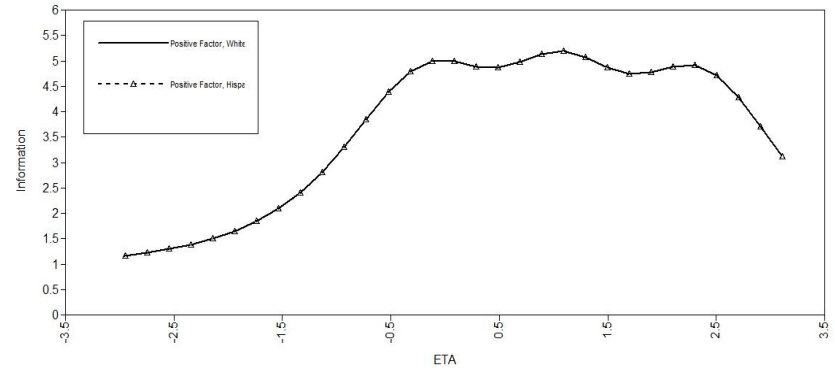b) ICC for Item 9 – *You thought your life had been a failure.*

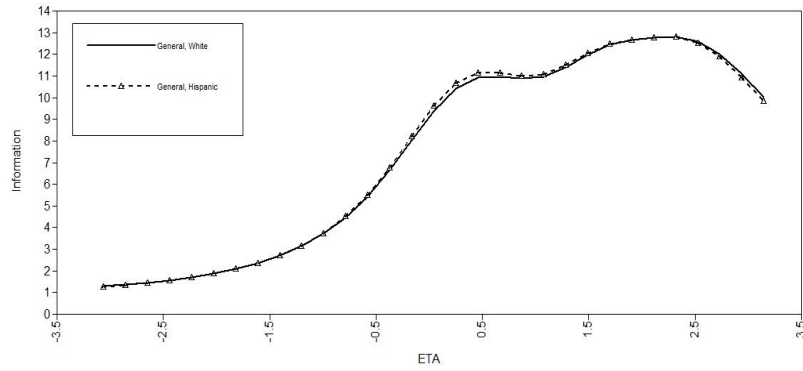c) ICC for Item 15 - *People were unfriendly to you.*

Figure 2. Item Characteristic Curves for Latinos (dashed lines) and Whites (solid lines) for three items that showed largest DIF in Model 2.
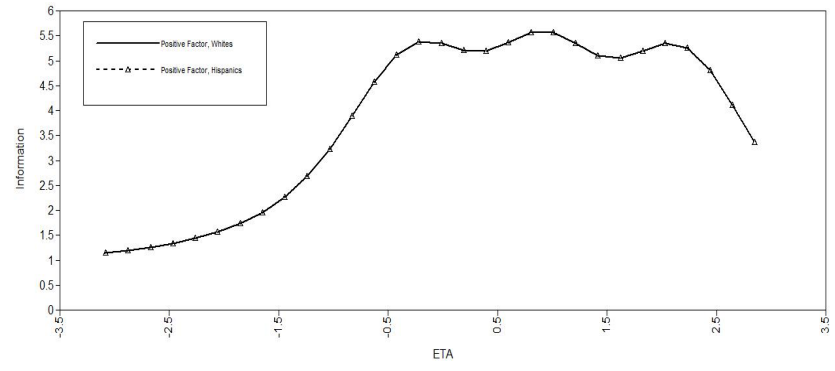
a) Model 1 – Negative Factor

b) Model 1 – Positive Factor

c) Model 2 – Negative Factor

d) Model 2– Positive Factor

Figure 3. Total Information Curves for Models 1 and 2.

APPENDIX 1: SUMMARY OF CES-D CHANGES IN ADD HEALTH

Original CES-D	Add Health Changes
During the Past Week	**During the Past Week***
0 – *Rarely or none of the time (less than 1 day)*; 1 – *Some or a little of the time (1-2 days)*; 2 – *Occasionally or a moderate amount of time (3-4 days)*; 3 – *Most or all of the time (5-7 days)*	0 – *Never or rarely*; 1 – *Sometimes*; 2 – *A lot of the time*; 3 – *Most of the time or all of the time*
1. I was bothered by things that usually don't bother me.	*You* were bothered by things that don't usually bother *you*.
2. I did not feel like eating; my appetite was poor.	*You* didn't feel like eating, *your* appetite was poor.
3. I felt that I could not shake off the blues even with help from my family or friends.	*You* felt that you could not shake off the blues, even with the help from *your* family and friends.
4. I felt I was just as good as other people.	*You* felt you were just as good as other people.^
5. I had trouble keeping my mind on what I was doing.	*You* had trouble keeping your mind on what you were doing.
6. I felt depressed.	*You* felt depressed.
7. I felt that everything I did was an effort.	*You felt that you were too tired to do things.*
8. I felt hopeful about the future.	*You* felt hopeful about the future.^
9. I thought my life had been a failure.	*You* thought *your* life had been a failure.
10. I felt fearful.	*You* felt fearful.
11. My sleep was restless.	*In the past 12 months, how often have you had trouble falling asleep or staying asleep?* 0 – *never*; 1 – *just a few times*; 2 – *about once a week*; 3 – *almost every day*; 4 – *every day*
12. I was happy.	*You were* happy.^
13. I talked less than usual.	*You* talked less than usual.
14. I felt lonely.	*You* felt lonely.
15. People were unfriendly.	People were unfriendly *to you*.
16. I enjoyed life.	*You* enjoyed life.^
17. I had crying spells.	*In the past 12 months, how often have you cried frequently?* 0 – *never*; 1 – *just a few times*; 2 – *about once a week*; 3 – *almost every day*; 4 – *every day*
18. I felt sad.	*You* felt sad.
19. I felt that people dislike me.	*You* felt that people disliked you.
20. I could not get "going."	*It was hard to get started doing things.*

Note: Changes to the wording are italicized. *Two questions asked about the past 12 months (items 11 and 17). ^Positive affect items, which are reverse scored.

REFERENCES

Alegria, M. (2009). The challenge of acculturation measures: what are we missing? A commentary on Thomson & Hoffman-Goetz. *Social Science & Medicine (1982)*, *69*(7), 996–998. https://doi.org/10.1016/j.socscimed.2009.07.006 [doi]

Alegría, M., Canino, G., Shrout, P. E., Woo, M., Duan, N., Vila, D., … Meng, X. (2008). Prevalence of mental illness in immigrant and non-immigrant U.S. groups. *American Journal of Psychiatry*, *165*(3), 359–369. https://doi.org/10.1176/appi.ajp.2007.07040704

Alegria, M., & Woo, M. (2009). Conceptual issues in Latino mental health. In F. A. Villaruel, G. Carlo, J. M. Grau, M. Azmitia, N. J. Cabrera, & T. J. Chahin (Eds.), *Handbook of U.S. Latino psychology: Developmental and community-based perspectives.* (pp. 15–30).

Allen, L., & Astuto, J. (2009). Depression among racially, ethnically, and culturally diverse adolescents. In S. Nolen-Hoeksema, L. M. Hilt, S. Nolen-Hoeksema (Ed), & L. M. Hilt (Ed) (Eds.), *Handbook of depression in adolescents.* (pp. 75–110). New York, NY, US: Routledge/Taylor & Francis Group. Retrieved from https://auth.lib.unc.edu/ezproxy_auth.php?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2009-01057-004&site=ehost-live&scope=site

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5). *Diagnostic and Statistical Manual of Mental Disorders 4th Edition TR.*, 280. https://doi.org/10.1176/appi.books.9780890425596.744053

Angold, A. (2002). Diagnostic interviews with parents and children. In M. Rutter & E. Taylor (Eds.), *Child and adolescent psychiatry: Modern approaches* (4th ed., pp. 32–51). Oxford, UK: Blackwell Scientific.

Arciniega, G. M., Anderson, T. C., Tovar-Blank, Z. G., & Tracey, T. J. G. (2008). Toward a fuller conception of Machismo: Development of a traditional Machismo and Caballerismo Scale. *Journal of Counseling Psychology*, *55*(1), 19–33. https://doi.org/10.1037/0022-0167.55.1.19

Bas-Sarmiento, P., Saucedo-Moreno, M. J., Fernández-Gutiérrez, M., & Poza-Méndez, M. (2017). Mental Health in Immigrants Versus Native Population: A Systematic Review of the Literature. *Archives of Psychiatric Nursing*, *31*(1), 111–121. https://doi.org/10.1016/j.apnu.2016.07.014

Bauer, D. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*, *22*(3), 507–526. https://doi.org/10.1037/met0000077

Bauer, D., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*, *14*(2), 101–125. https://doi.org/10.1037/a0015583

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine*, *25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Berry, J. W. (2003). Conceptual approaches to acculturation. *Acculturation: Advances in Theory, Measurement, and Applied Research*, 17–37. https://doi.org/10.1037/10472-004

Birmaher, B., Arbelaez, C., & Brent, D. (2002). Course and outcome of child and adolescent major depressive disorder. *Child and Adolescent Psychiatric Clinics of North America*. https://doi.org/10.1016/S1056-4993(02)00011-1

Bollen, K. A. (1989). *Structural equations with latent variables*. *Psychometrika*. https://doi.org/10.1007/BF02293905

Booth-Kewley, S., Rosenfeld, P., & Edwards, J. E. (1992). Impression management and self-deceptive enhancement among hispanic and non-hispanic white navy recruits. *Journal of Social Psychology*. https://doi.org/10.1080/00224545.1992.9924707

Borsboom, D. (2006). When Does Measurement Invariance Matter? *Medical Care*, *44*(Suppl 3), S176–S181. https://doi.org/10.1097/01.mlr.0000245143.08679.cc

Boutin-Foster, C. (2008). An item-level analysis of the Center for Epidemiologic Studies Depression Scale (CES-D) by race and ethnicity in patients with coronary artery disease. *International Journal of Geriatric Psychiatry*, 2.

Braveman, P. A., Cubbin, C., Egerter, S., Chideya, S., Marchi, K. S., Metzler, M., & Posner, S. (2005). Socioeconomic status in health research: One size does not fit all. *Journal of the American Medical Association*, *294*(22), 2879–2888. https://doi.org/10.1001/jama.294.22.2879

Cabassa, L. J. (2003). Measuring Acculturation : Where We Are and Where We Need to Go. *Hispanic Journal of Behavioral Sciences*, *25*(2), 127–146. https://doi.org/10.1177/0739986303253626

Canino, I. A., Rubio-Stipec, M., Canino, G., & Escobar, J. I. (1992). Functional somatic symptoms: a cross-ethnic comparison. *The American Journal of Orthopsychiatry*, *62*(4), 605–612. https://doi.org/10.1037/h0079376

Cauce, A. M., & Domenech-Rodriguez, M. (2002). Latino Families: Myths and Realities. In J. M. Contreras, K. A. Kerns, & A. M. Neal-Barnett (Eds.), *Latino Children and Families in the United States: Current Research and Future Directions* (pp. 3–25).

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*(5), 1005–1018. https://doi.org/10.1037/a0013193

Chen, P., & Chantala, K. (2013). *Guidelines for analyzing Add Health data*.

Choi, H., & Park, C. G. (2006). Understanding adolescent depression in ethnocultural context: Updated with Empirical Findings. *Advances in Nursing Science*, *29*(5), E1–E12.

Cole, V. T., Gottfredson, N. C., Giordano, M., & Janssen, T. (2018). Automated fitting of moderated nonlinear factor analysis (MNLFA) through the Mplus program.

Conger, R. D., Ge, X., Elder, G. H., Lorenz, F. O., & Simons, R. L. (1994). Economic Stress, Coercive Family Process, and Developmental Problems of Adolescents. *Child Development*. https://doi.org/10.1111/j.1467-8624.1994.tb00768.x

Costello, E. J., Egger, H., & Angold, A. (2005). 10-Year research update review: The epidemiology of child and adolescent psychiatric disorders: I. Methods and public health burden. *Journal of the American Academy of Child and Adolescent Psychiatry*. https://doi.org/10.1097/01.chi.0000172552.41596.6f

Crane, P. K., van Belle, G., & Larson, E. B. (2004). Test bias in a cognitive test: Differential item functioning in the CASI. *Statistics in Medicine*. https://doi.org/10.1002/sim.1713

Crockett, L. J., Randall, B. A., Shen, Y.-L., Russell, S. T., & Driscoll, A. K. (2005). Measurement Equivalence of the Center for Epidemiological Studies Depression Scale for Latino and Anglo Adolescents: A National Study. *Journal of Consulting and Clinical Psychology*, *73*(1), 47–58. https://doi.org/10.1037/0022-006X.73.1.47

Cuellar, I., Arnold, B., & Maldonado, R. (1995). Acculturation Rating Scale for Mexican Americans-II: A Revision of the Original ARSMA Scale. *Hispanic Journal of Behavioral Sciences*. https://doi.org/10.1177/07399863950173001

Curran, P. J., Cole, V., Bauer, D. J., Hussong, A. M., & Gottfredson, N. (2016). Improving Factor Score Estimation Through the Use of Observed Background Characteristics. *Structural Equation Modeling*, *23*(6), 827–844. https://doi.org/10.1080/10705511.2016.1220839

Curran, P. J., Cole, V. T., Bauer, D. J., Rothenberg, W. A., & Hussong, A. M. (2018). Recovering Predictor–Criterion Relations Using Covariate-Informed Factor Score Estimates. *Structural Equation Modeling*, 1–16. https://doi.org/10.1080/10705511.2018.1473773

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*(2), 81–100. https://doi.org/10.1037/a0015914

Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A., Chassin, L., … Zucker, R. (2014). A Moderated Nonlinear Factor Model for the Development of Commensurate Measures in Integrative Data Analysis. *Multivariate Behavioral Research*, *49*(3), 214–231. https://doi.org/10.1080/00273171.2014.889594

Cyranowski, J. M., Frank, E., Young, E., & Shear, M. K. (2000). Adolescent onset of the gender difference in lifetime rates of major depression: a theoretical model. *Arch Gen Psychiatry*, *57*(1), 21–27. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10632229

De Ayala, R. J. (2008). The theory and practice of item response theory. *Methodology in the Social Sciences*, 448. Retrieved from http://books.google.com/books?id=-k36zbOBa28C

Edelbrock, C., Costello, A. J., Dulcan, M. K., Kalas, R., & Conover, N. C. (1985). Age differences in the reliability of the psychiatric interview of the child. *Child Development*, *56*(1), 265–275. https://doi.org/10.2307/1130193

Edwards, M. C., Cheavens, J. S., Heiy, J. E., & Cukrowicz, K. C. (2010). A Reexamination of the Factor Structure of the Center for Epidemiologic Studies Depression Scale: Is a One-Factor Model Plausible? *Psychological Assessment*, *22*(3), 711–715. https://doi.org/10.1037/a0019917

Ennis, S. R., Rios-Vargas, M., & Albert, N. G. (2011). The Hispanic Population : 2010. *201*, (May), 16. https://doi.org/10.1016/j.mhpa.2012.10.005

Epstein, R. M., Quill, T. E., & McWhinney, I. R. (1999). Somatization reconsidered: incorporating the patient's experience of illness. *Arch Intern Med*, *159*(3), 215–222. https://doi.org/10.1001/archinte.159.3.215

Escobar, J. I., Burnam, M. A., Karno, M., Forsythe, A., & Golding, J. M. (1987). Somatization in the Community. *Archives of General Psychiatry*, *44*(8), 713. https://doi.org/10.1001/archpsyc.1987.01800200039006

Ferraz, M. B. (1997). Cross cultural adaptation of questionnaires: what is it and when should it be performed? *The Journal of Rheumatology*, *24*(11), 2066.

Fragoso, J. M., & Kashubeck, S. (2000). Machismo, gender role conflict, and mental health in Mexican American men. *Psychology of Men & Masculinity*, *1*(2), 87–97. https://doi.org/10.1037/1524-9220.1.2.87

Garber, J., & Horowitz, J. L. (2002). Depression in children. In I. H. Gotlib & C. L. Hammen (Eds.), *Handbook of depression* (pp. 510–540). New York, NY, US: Guilford Press.

Garber, J., & Kaminski, K. M. (2000). Laboratory and performance-based measures of depression in children and adolescents. *Journal of Clinical Child Psychology*, *29*(4), 509–525. https://doi.org/10.1207/S15374424JCCP2904_5

Gonzales, N. A., Knight, G. P., Morgan-Lopez, A. A., Saenz, D., & Sirolli, A. (2002). Acculturation and the mental health of Latino youths: An integration and critique of the literature. In *Latino children and families in the United States: Current research and future directions*.

Gonzalez-Barrera, A. (2015). Mestizo'and 'mulatto': Mixed-race identities among US Hispanics. *Pew Center for Research*, *10*.

Goodman, E., Slap, G. B., & Huang, B. (2003). The Public Health Impact of Socioeconomic Status on Adolescent Depression and Obesity. *American Journal of Public Health*. https://doi.org/10.2105/AJPH.93.11.1844

Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: a pilot study on the validity of the self-report version. *Eur Child Adolesc Psychiatry*, *7*(3), 125–130. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9826298

Gottfredson, N. C., Cole, V. T., Giordano, M. L., Bauer, D. J., Hussong, A. M., & Ennett, S. T. (2018). Simplifying the implementation of modern scale scoring methods with an automated R package: Automated moderated nonlinear factor analysis (aMNLFA). *Addictive Behaviors*. https://doi.org/10.1016/j.addbeh.2018.10.031

Greenman, E., & Xie, Y. (2008). Is assimilation theory dead? The effect of assimilation on adolescent well-being. *Social Science Research*. https://doi.org/10.1016/j.ssresearch.2007.07.003

Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology*, *46*(12), 1417–1432. https://doi.org/10.1016/0895-4356(93)90142-N

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. *Structural Equation Modeling*. https://doi.org/10.1080/10705511.2017.1402334

Hambleton, R. K., & Kanjee, A. (1995). Increasing the validity of cross-cultural assessments: Use of improved methods for test adaptations. *European Journal of Psychological Assessment*, *11*(3), 147–157. https://doi.org/10.1027/1015-5759.11.3.147

Hammen, C., & Rudolph, K. D. (2003). Childhood mood disorders. In *Child psychopathology, 2nd ed.* (pp. 233–278). New York, NY, US: Guilford Press.

Hankin, B. L., Abramson, L. Y., Moffitt, T. E., Angell, K. E., Silva, P. A., & McGee, R. (1998). Development of depression from preadolescence to young adulthood: Emerging gender differences in a 10-year longitudinal study. *Journal of Abnormal Psychology*. https://doi.org/10.1037/0021-843X.107.1.128

Harris, K. M., & Udry, J. R. (2008). *National longitudinal study of adolescent health (add health), 1994-2008*. Inter-university Consortium for Political and Social Research.

Hilt, L., & Nolen-Hoeksema, S. (2009). The emergence of gender differences in depression in adolescence. In S. Nolen-Hoeksema & L. M. Hilt (Eds.), *Handbook of depression in adolescents* (pp. 111–135). New York: Routledge.

Holland, P. W., & Wainer, H. (1993). *Differential item functioning. Differential item functioning*. Retrieved from http://www.redi-bw.de/db/ebsco.php/search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1993-97193-000&site=ehost-live

Hopwood, C. J., Flato, C. G., Ambwani, S., Garland, B. H., & Morey, L. C. (2009). A comparison of Latino and Anglo socially desirable responding. *Journal of Clinical Psychology*, *65*(7), 769–780. https://doi.org/10.1002/jclp.20584

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*. https://doi.org/10.1080/03610739208253916

Horn, J. L., McArdle, J., & Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *Southern Psychologist*.

Hovey, J. D., & King, C. a. (1996). Acculturative stress, depression, and suicidal ideation among immigrant and second-generation Latino adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*. https://doi.org/10.1097/00004583-199609000-00016

Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*. https://doi.org/10.1177/0022022189203004

Hurtado, A. (1992). *Redefining California: Latino social engagement in a multicultural society*. Chicano Studies Research Center.

Iwata, N., & Buka, S. (2002). Race/ethnicity and depressive symptoms: a cross-cultural/ethnic comparison among university students in East Asia, North and South America. *Soc Sci Med*, *55*(12), 2243–2252. https://doi.org/10.1016/S0277-9536(02)00003-5

Iwata, N., Turner, R. J., & Lloyd, D. A. (2002). Race/ethnicity and depressive symptoms in community-dwelling young adults: A differential item functioning analysis. *Psychiatry Research*, *110*(3), 281–289. https://doi.org/10.1016/S0165-1781(02)00102-6

Jensen, B. J., & Haynes, S. N. (1986). Self-report questionnaires and inventories. In *Handbook of behavioral assessment* (Vol. 88, p. 150). John Wiley & Sons Inc.

Johnson, T. P., & Van de Vijver, F. J. R. (2002). Social desirability in cross-cultural research. *Cross-Cultural Survey Methods*. https://doi.org/10.4135/9781412963947.n537

Katragadda, C. P., & Tidwell, R. (1998). Rural hispanic adolescents at risk for depressive symptoms. *Journal of Applied Social Psychology*.

Kessing, L. V., Hansen, M. G., & Andersen, P. K. (2004). Course of illness in depressive and bipolar disorders: Naturalistic study, 1994-1999. *British Journal of Psychiatry*, *185*(NOV.), 372–377. https://doi.org/10.1192/bjp.185.5.372

Kessler, R. C., Avenevoli, S., & Ries Merikangas, K. (2001). Mood disorders in children and adolescents: An epidemiologic perspective. *Biological Psychiatry*. https://doi.org/10.1016/S0006-3223(01)01129-5

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., … Wang, P. S. (2003). The Epidemiology of Major Depressive Disorder: Results from the National Comorbidity Survey Replication (NCS-R). *JAMA*, *289*(23), 3095–3105.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005a). Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593. https://doi.org/10.1001/archpsyc.62.6.593

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005b). Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*, *62*(6), 617–627. https://doi.org/10.1001/archpsyc.62.6.617

Kessler, R. C., Merikangas, K. R., & Wang, P. S. (2007). Prevalence, Comorbidity, and Service Utilization for Mood Disorders in the United States at the Beginning of the Twenty-first Century. *Annual Review of Clinical Psychology*, *3*(1), 137–158. https://doi.org/10.1146/annurev.clinpsy.3.022806.091444

Kim, G., Chiriboga, D. A., & Jang, Y. (2009). Cultural equivalence in depressive symptoms in older white, black, and mexican-American adults: Clinical investigations. *Journal of the American Geriatrics Society*, *57*(5), 790–796. https://doi.org/10.1111/j.1532-5415.2009.02188.x

Kim, G., Decoster, J., Huang, C.-H., & Chiriboga, D. A. (2011). Race/ethnicity and the factor structure of the Center for Epidemiologic Studies Depression Scale: a meta-analysis. *Cultural Diversity & Ethnic Minority Psychology*, *17*(4), 381–396. https://doi.org/10.1037/a0025434

Kirmayer, L. J. (2001). Cultural Variations in the Clinical Presentation of Depression and Anxiety: Implications for Diagnosis and Treatment. *J Clin Psychiatry*, *6262*, 22–28.

Kirmayer, L. J., Robbins, J. M., Dworkind, M., & Yaffe, M. J. (1993). Somatization and the Recognition of Depression and Anxiety in Primary Care. *The American Journal of Psychiatry*, *150*(5), 734–741. https://doi.org/10.1136/bmj.319.7204.254b

Knight, G. P., Tein, J.-Y., Prost, J. H., & Gonzales, N. A. (2002). Measurement Equivalence and Research on Latino children and families: The importance of culturally informed theory. In J. M. Contreras, K. A. Kerns, A. M. Neal-Barnett, J. M. Contreras (Ed), K. A. Kerns (Ed), & A. M. Neal-Barnett (Ed) (Eds.), *Latino children and families in the United States: Current research and future directions.* (pp. 181–201). Westport, CT, US: Praeger Publishers/Greenwood Publishing Group. Retrieved from https://auth.lib.unc.edu/ezproxy_auth.php?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2007-04486-008&site=ehost-live&scope=site

Kolody, B., Vega, W., Meinhardt, K., & Bensussen, G. (1986). The correspondence of health complaints and depressive symptoms among Anglos and Mexican-Americans. *The Journal of Nervous and Mental Disease*, *174*(4), 221–228. https://doi.org/10.1097/00005053-198604000-00005

Kopper, B. a, & Epperson, D. L. (1996). The experience and expression of anger: Relationships with gender, gender role socialization, depression, and mental health functioning. *Journal of Counseling Psychology*, *43*(2), 158–165. https://doi.org/10.1037/0022-0167.43.2.158

Kroenke, K., Jackson, J. L., Chamberlin, J., Kroenke, K., Jackson, J. L., & Chamberlin, J. (1997). Depressive and Anxiety Disorders in Patients Presenting With Physical Complaints. *The American Journal of Medicine*, *103*(5), 339–347. https://doi.org/S0002-9343(97)00241-6 [pii]

Kubik, M. Y., Lytle, L. A., Birnbaum, A. S., Murray, D. M., & Perry, C. L. (2003). Prevalence and Correlates of Depressive Symptoms in Young Adolescents. *American Journal of Health Behavior*, *27*(5), 546–553. https://doi.org/10.5993/AJHB.27.5.6

Labouvie, E., & Ruetsch, C. (1995). Testing for Equivalence of Measurement Scales: Simple Structure and Metric Invariance Reconsidered. *Multivariate Behavioral Research*. https://doi.org/10.1207/s15327906mbr3001_4

Levitt, P., DeWind, J., & Vertovec, S. (2006). International Perspectives on Transnational Migration: An Introduction. *International Migration Review*. https://doi.org/10.1111/j.1747-7379.2003.tb00150.x

Lewinsohn, P. M., & Essau, C. A. (2002). Depression in adolescents. In *Handbook of depression.* (pp. 541–559). New York, NY, US: Guilford Press.

Lewis-Fernández, R., Das, A. K., Alfonso, C., Weissman, M. M., & Olfson, M. (2005). Depression in US Hispanics: diagnostic and management considerations in family practice. *The Journal of the American Board of Family Practice / American Board of Family Practice*, *18*(4), 282–296. https://doi.org/10.3122/jabfm.18.4.282

Liang, J., Tran, T. V., Krause, N., & Markides, K. S. (1989). Generational differences in the structure of the CES-D scale in Mexican Americans. *Journals of Gerontology*. https://doi.org/10.1093/geronj/44.3.S110

MacIntosh, R. C., & Strickland, O. J. (2010). Differential item responses on CES-D inventory: A comparison of elderly Hispanics and non-Hispanic Whites in the United States and item usage by elderly Hispanics across time. *Aging & Mental Health*, *14*(5), 556–564. https://doi.org/10.1080/13607860903421045

Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme Response Style and Acquiescence Among Hispanics. *Hispanic Journal of Behavioral Sciences*, *9*(2), 498–509. https://doi.org/10.1177/07399863870092005

Marotta, S. A., & Garcia, J. G. (2003). Latinos in the United States in 2000. *Hispanic Journal of Behavioral Sciences*. https://doi.org/10.1177/0739986303251693

McArdle, J. J., Johnson, R. C., Hishinuma, E. S., Miyamoto, R. H., & Andrade, N. N. (2001). Structural equation modeling of group differences in CES-D ratings of native Hawaiian and non-Hawaiian high school students. *Journal of Adolescent Research*. https://doi.org/10.1177/0743558401162002

McCabe, B. E., Vermeesch, A. L., Hall, R. F., Peragallo, N. P., & Mitrani, V. B. (2011). Acculturation and the Center for Epidemiological Studies-Depression Scale for Hispanic Women. *Nursing Research*, *60*(4), 270–275. https://doi.org/10.1097/NNR.0b013e318221b8dc

Mchorney, C. A., & Fleishman, J. A. (2010). Assessing and Understanding Measurement Equivalence in Health Outcome Measures. *Medical Care*, *44*(11), 205–210.

Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics*, *7*(2), 105–118.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543.

Miller, T. Q., Markides, K. S., & Black, S. A. (1997). The factor structure of the CES-D in two surveys of elderly Mexican Americans. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*. https://doi.org/10.1093/geronb/52B.5.S259

Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. New York: Routledge.

Millsap, R. E., & Everson, H. T. (1993). Methodology Review: Statistical Approaches for Assessing Measurement Bias. *Applied Psychological Measurement*, *17*(4), 297–334. https://doi.org/10.1177/014662169301700401

Motel, S., & Patten, E. (2012). The 10 Largest Hispanic Origin Groups: Characteristics, Rankings, Top Counties. *Pew Research*, (202), 1–5. Retrieved from http://www.pewhispanic.org/2012/06/27/the-10-largest-hispanic-origin-groups-characteristics-rankings-top-counties/%5Cnpapers2://publication/uuid/AA125F86-1DE1-4479-87B1-BA38359A8E6B

Muthén, L., & Muthén, B. (2007). *Mplus user's guide (version 7). Los Angeles: Author*.

Nguyen, H. T., Clark, M., & Ruiz, R. J. (2007). Effects of acculturation on the reporting of depressive symptoms among Hispanic pregnant women. *Nursing Research*, *56*(3), 217–223. https://doi.org/10.1097/01.NNR.0000270027.97983.96

Niemann, Y. F. (2004). Stereotypes of Chicanas and Chicanos: Impact on family functioning, individual expectations, goals, and behavior. In R. J. Velasquez, L. M. Arellano, & B. W. McNell (Eds.), *The handbook of Chicana/o psychology and mental health* (pp. 61–82). Mahwah, NJ: Lawrence Erlbaum Assoc Inc.

Nock, M. K., & Kessler, R. C. (2006). Prevalence of and risk factors for suicide attempts versus suicide gestures: Analysis of the National Comorbidity Survey. *Journal of Abnormal Psychology*, *115*(3), 616–623. https://doi.org/10.1037/0021-843X.115.3.616

Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *J Abnorm Psychol*, *109*(3), 504–511. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11016119

Oetting, E. R., Donnermeyer, J. F., Trimble, J. E., & Beauvais, F. (1998). Primary socialization theory: Culture, ethnicity, and cultural identification. The links between culture and substance use. IV. *Substance Use and Misuse*. https://doi.org/10.3109/10826089809069817

Ojeda, L., Rosales, R., & Good, G. E. (2008). Socioeconomic status and cultural predictors of male role attitudes among Mexican American men: Son más machos? *Psychology of Men & Masculinity*, *9*(3), 133–138. https://doi.org/10.1037/1524-9220.9.3.133

Paxton, R. J., Valois, R. F., Watkins, K. W., Huebner, E. S., & Drane, J. W. (2007). Sociodemographic differences in depressed mood: results from a nationally representative sample of high school adolescents. *The Journal of School Health*, *77*(4), 180–186. https://doi.org/10.1111/j.1746-1561.2007.00189.x

Perreira, K. M., Deeb-Sossa, N., Harris, K. M., & Bollen, K. (2005). What Are We Measuring? An Evaluation of the CES-D Across Race/Ethnicity and Immigrant Generation. *Social Forces*, *83*(4), 1567–1602. https://doi.org/10.1353/sof.2005.0077

Phinney, J. S., & Flores, J. (2002). "Unpackaging" acculturation: Aspects of acculturation as predictors of traditional sex role attitudes. *Journal of Cross-Cultural Psychology*, *33*(3), 320–331. https://doi.org/10.1177/0022022102033003007

Pina-Watson, B., Castillo, L. G., Ojeda, L., & Rodriguez, K. M. (2013). Parent conflict as a mediator between marianismo beliefs and depressive symptoms for mexican american college women. *Journal of American College Health*, *61*(8), 491–496. https://doi.org/10.1080/07448481.2013.838567

Pina, A. A., Gonzales, N. A., Holly, L. E., Zerr, A. A., & Wynne, H. (2013). Toward evidence-based clinical assessment of ethnic minority youth. In B. D. McLeod, A. Jensen-Doss, T. H. Ollendick, B. D. McLeod (Ed), A. Jensen-Doss (Ed), & T. H. Ollendick (Ed) (Eds.), *Diagnostic and behavioral assessment in children and adolescents: A clinical guide.* (pp. 348–376). New York, NY, US: Guilford Press. Retrieved from https://auth.lib.unc.edu/ezproxy_auth.php?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2013-27413-013&site=ehost-live&scope=site

Pina, A. A., Little, M., Knight, G. P., & Silverman, W. K. (2009). Cross-ethnic measurement equivalence of the RCMAS in Latino and White youth with anxiety disorders. *Journal of Personality Assessment*, *91*(1), 58–61. https://doi.org/10.1080/00223890802484183

Quiñones-Rosado, R. (2002). Empowerment, Latino identity, and the process of transformation: Writing for people in the struggle against racism. Caguas, PR: Institute for Latino Empowerment.

Radloff, L. S. (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, *1*(3), 385–401. https://doi.org/10.1177/014662167700100306

Ramos, J. M., & Magana, C. G. (2008). *Depression in Latinos: Assessment, Treatment, and Prevention. Depression in Latinos: Assessment, Treatment, and Prevention*. https://doi.org/10.1007/978-0-387-78512-7

Richters, J., & Pellegrini, D. (1989). Depressed mothers' judgments about their children: An examination of the depression-distortion hypothesis. *Child Development*, *60*(5), 1068–1075. https://doi.org/10.2307/1130780

Roberts, R. E., & Sobhan, M. (1992). Symptoms of depression in adolescence: A comparison of Anglo, African, and Hispanic Americans. *Journal of Youth and Adolescence*, *21*(6), 639–651. https://doi.org/10.1007/BF01538736

Rogler, L. H., Cortes, D. E., & Malgady, R. G. (1991). Acculturation and mental health status among hispanics: Convergence and new directions for research. *American Psychologist*. https://doi.org/10.1037/0003-066X.46.6.585

Ross, C. E., & Mirowsky, J. (2006). Socially-Desirable Response and Acquiescence in a Cross-Cultural Survey of Mental Health. *Journal of Health and Social Behavior*. https://doi.org/10.2307/2136668

Rumbaut, R. G. (1996). The new Californians: Assessing the educational progress of children of immigrants. *CPS Brief*.

Shavers, V. L. (2007). Measurement of socioeconomic status in health disparities research. *Journal of the National Medical Association*, *99*(9), 1013–1023. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17913111%0Ahttp://www.pubmedcentral.nih.gov/

articlerender.fcgi?artid=PMC2575866

Shih, J. H., Eberhart, N. K., Hammen, C. L., & Brennan, P. A. (2006). Differential exposure and reactivity to interpersonal stress predict sex differences in adolescent depression. *Journal of Clinical Child and Adolescent Psychology*. https://doi.org/10.1207/s15374424jccp3501_9

Simon, G. E., VonKorff, M., Piccinelli, M., Fullerton, C., & Ormel, J. (1999). An international study of the relation between somatic symptoms and depression. *The New England Journal of Medicine*, *341*(18), 1329–1335. https://doi.org/10.1056/NEJM199910283411801

Spitzer, R. L. (1983). Psychiatric diagnosis: are clinicians still necessary? *Comprehensive Psychiatry*.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2004). Examining the Effects of Differential Item (Functioning and Differential) Test Functioning on Selection Decisions: When Are Statistically Significant Effects Practically Important? *Journal of Applied Psychology*, *89*(3), 497–508. https://doi.org/10.1037/0021-9010.89.3.497

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*, *91*(6), 1292–1306. https://doi.org/10.1037/0021-9010.91.6.1292

Steinberg, L., & Thissen, D. (2006). Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. *Psychological Methods*. https://doi.org/10.1037/1082-989X.11.4.402

Stewart, T. D., & Reed, M. B. (2015). Lifetime nonmedical use of prescription medications and socioeconomic status among young adults in the United States. *American Journal of Drug and Alcohol Abuse*, *41*(5), 458–464. https://doi.org/10.3109/00952990.2015.1060242

Stricker, L. (1982). Identifying Test Items That Perform Differentially in Population Subgroups: A Partial Correlation Index. *Applied Psychological Measurement*, *6*(3), 261–273. https://doi.org/10.1177/014662168200600302

Stricker, L. (1984). The stability of a partial correlation index for identifying items that perform differentially in subgroups. *Educational and Psychological Measurement*, *44*(4), 831–837.

Stroup-Benham, C. A., Lawrence, R. H., & Treviño, F. M. (1992). CES-D factor structure among Mexican American and Puerto Rican women from single- and couple-headed households. *Hispanic Journal of Behavioral Sciences*, *14*(3), 310–326. https://doi.org/10.1177/07399863920143002

Syzdek, M. R., & Addis, M. E. (2010). Adherence to masculine norms and attributional processes predict depressive symptoms in recently unemployed men. *Cognitive Therapy and Research*, *34*(6), 533–543. https://doi.org/10.1007/s10608-009-9290-6

Teresi, J. A., Ramirez, M., Lai, J.-S., & Silver, S. (2008). Occurrences and sources of Differential Item Functioning (DIF) in patient-reported outcome measures: Description of DIF methods, and review of measures of depression, quality of life and general health. *Psychology Science Quarterly*, *50*(4), 538. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20165561%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2822361

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics*, *27*(1), 77–83.

U.S. Census Bureau. (2015). Annual estimates of the resident population by sex, age, race, and Hispanic Origin for the United States, states, and counties: April 1, 2010 to July 1, 2014. *American FactFinder*, 2015–2016. Retrieved from http://factfinder.census.gov/bkmk/table/1.0/en/PEP/2014/PEPASR6H/0100000US%5Cnfiles/331/productview.html

Uebelacker, L. A., Strong, D., Weinstock, L. M., & Miller, I. W. (2009). Use of item response theory to understand differential functioning of DSM-IV major depression symptoms by race, ethnicity and gender. *Psychological Medicine*, *39*(04), 591. https://doi.org/10.1017/S0033291708003875

Umaña-Taylor, A. J., & Updegraff, K. A. (2007). Latino adolescents' mental health: Exploring the interrelations among discrimination, ethnic identity, cultural orientation, self-esteem, and depressive symptoms. *Journal of Adolescence*, *30*(4), 549–567. https://doi.org/10.1016/j.adolescence.2006.08.002

Van Beek, Y., Hessen, D. J., Hutteman, R., Verhulp, E. E., & Van Leuven, M. (2012). Age and gender differences in depression across adolescence: Real or "bias"? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *53*(9), 973–985. https://doi.org/10.1111/j.1469-7610.2012.02553.x

Vandenberg, R. J., & Lance, C. E. (2007). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Varela, R. E., Weems, C. F., Berman, S. L., Hensley, L., & De Bernal, M. C. R. (2007). Internalizing symptoms in latinos: The role of anxiety sensitivity. *Journal of Youth and Adolescence*, *36*(4), 429–440. https://doi.org/10.1007/s10964-007-9168-4

Vaughn-Coaxum, R. A., Mair, P., & Weisz, J. R. (2016). Racial/Ethnic Differences in Youth Depression Indicators. *Clinical Psychological Science*, *4*(2), 239–253. https://doi.org/10.1177/2167702615591768

Vazquez-Nuttall, E., Romero-Garcia, I., & de Leon, B. (1987). Sex roles and perceptions of femininity and masculinity of hispanic women: A Review of the Literature. *Psychology of Women Quarterly*, *11*(4), 409–425. https://doi.org/10.1111/j.1471-6402.1987.tb00915.x

Vega, W. A., & Rumbaut, R. G. (1991). Ethnic minorities and Mental Health. *Annual Review of Sociology*, *17*, 351–383. https://doi.org/10.1146/annurev.so.17.080191.002031

Vega, W. A., Zimmerman, R., Gil, A., Warheit, G. J., & Apospori, E. (1997). Acculturation strain theory: Its application in explaining drug use behavior among Cuban and other Hispanic youth. *Substance Use and Misuse*. https://doi.org/10.3109/10826089709035608

Viruell-Fuentes, E. A. (2006). "My heart is always there": The transnational practices of first-generation Mexican immigrant and second-generation Mexican American women. *Identities: Global Studies in Culture and Power*. https://doi.org/10.1080/10702890600838076

Waslick, B. D., Kandel, R., & Kakouros, A. (2002). Depression in children and adolescents: An overview. In D. Shaffer & B. D. Waslick (Eds.), *The many faces of depression in children and adolescents.* (pp. 1–36). American Psychological Association.

Weissman, M. M., Orvaschel, H., & Padian, N. (1980). Children's symptom and social functioning self-report scales comparison of mothers' and children's reports. *Journal of Nervous and Mental Disease*. https://doi.org/10.1097/00005053-198012000-00005

Wells, K. B., Golding, J. M., Hough, R. L., Burnam, M. a, & Karno, M. (1989). Acculturation and the probability of use of health services by Mexican Americans. *Health Services Research*, *24*(2), 237–257.

Yi, J. Y., Chen, Y., Hussong, A. M., & Daughters, S. B. (2016). Acculturation and Substance Use among Hispanic-Latino and Asian Adolescents. *Current Psychiatry Reviews*. https://doi.org/10.2174/1573400512666160216234716

Zich, J. M., Attkisson, C. C., & Greenfield, T. K. (1990). Screening for depression in primary care clinics: the CES-D and the BDI. *International Journal of Psychiatry in Medicine*, *20*(3), 259–277. https://doi.org/10.2190/LYKR-7VHP-YJEM-MKM2